

# ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI IDENTITAS KEPENDUDUKAN DIGITAL DENGAN METODE NAIVE BAYES DAN K-NEAREST

Dany Pratmanto<sup>1</sup>, Fabriyan Fandi Dwi Imaniawan<sup>2</sup>, Vadlya Maarif<sup>3</sup>

<sup>1,3</sup> Program Studi Teknologi Komputer, Fakultas Teknologi Informasi, Universitas Bina Sarana  
Informatika,

<sup>2</sup> Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Bina Sarana  
Informatika,

Jl. Kramat Raya No.98, Senen, Jakarta Pusat, Indonesia

E-mail: <sup>1</sup>dany.dto@bsi.ac.id, <sup>2</sup>fabriyan.fbf@bsi.ac.id, <sup>3</sup>vadlya.vlr@bsi.ac.id

## Abstrak

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap data ulasan pengguna aplikasi Identitas Kependudukan Digital (IKD) dengan membandingkan performa dua metode, yakni Naive Bayes dan K-Nearest Neighbor (KNN). Data ulasan diambil dari Google Play Store lalu dipreprocess melalui text cleaning, case folding, tokenization, filtering, stemming, serta removal stopword agar siap dianalisis. Selanjutnya dilakukan transformasi data tekstual menjadi representasi numerik melalui pembobotan TF-IDF dan ekstraksi fitur. Kedua metode machine learning kemudian diimplementasikan untuk melakukan klasifikasi ulasan ke dalam dua kelas sentimen, yaitu positif dan negatif. Hasil penelitian menunjukkan bahwa KNN secara signifikan mengungguli Naive Bayes dalam hal akurasi dan presisi klasifikasi sentimen pada data ulasan aplikasi IKD. KNN mampu mencapai akurasi rata-rata 82,85% dan presisi di atas 80% untuk kedua kelas sentimen. Walaupun demikian, masih terdapat peluang peningkatan performa dengan melakukan parameter tuning serta kombinasi dengan metode lain seperti SVM atau deep learning. Secara keseluruhan, penelitian ini berhasil membuktikan efektivitas penerapan Naive Bayes dan KNN untuk analisis sentimen, khususnya KNN yang menunjukkan hasil paling optimal. Hasil penelitian diharapkan dapat memberikan wawasan bermanfaat untuk evaluasi dan pengembangan aplikasi IKD agar lebih user-friendly dan sesuai dengan kebutuhan masyarakat Indonesia.

**Kata kunci**— sentiment analysis, text mining, Naive Bayes, KNN, sentiment classification

## Abstract

This study aims to conduct sentiment analysis on user review data of the Digital Identity Card (IKD) application by comparing the performance of two methods, Naive Bayes and K-Nearest Neighbor (KNN). The review data was collected from Google Play Store and then preprocessed through text cleaning, case folding, tokenization, filtering, stemming, and stopword removal to make it analysis-ready. The textual data was then transformed into numerical representations using TF-IDF weighting and feature extraction. The two machine learning methods were implemented to classify the reviews into two sentiment categories, positive and negative. The results showed that KNN significantly outperformed Naive Bayes in terms of accuracy and precision of sentiment classification on the IKD app review data. KNN achieved an average accuracy of 82.85% and precision above 80% for both sentiment classes. However, there are still opportunities to improve performance through parameter tuning and combining with other

*methods like SVM or deep learning. Overall, this study has proven the efficacy of applying Naive Bayes and KNN for sentiment analysis, especially KNN which demonstrated the most optimal results. The findings are expected to provide valuable insights for the evaluation and development of the IKD application to make it more user-friendly and aligned with the needs of Indonesian society.*

**Keywords**— *analisis sentimen, text mining, Naive Bayes, KNN, klasifikasi sentimen*

## 1. PENDAHULUAN

Kartu Tanda Penduduk elektronik (KTP-el) adalah identitas resmi warga negara Indonesia yang diterbitkan pemerintah secara elektronik. KTP-el pertama kali diperkenalkan pada tahun 2011 sebagai pengganti KTP konvensional berbasis kertas. Tujuan penerapan KTP-el adalah untuk meningkatkan kualitas pelayanan administrasi kependudukan, menjamin keabsahan identitas penduduk, dan mendorong reformasi birokrasi menuju tata kelola pemerintahan yang lebih baik.[1]

KTP-el memiliki sejumlah keunggulan dibandingkan KTP biasa. Data warga tersimpan secara terintegrasi dalam database elektronik sehingga memudahkan validasi dan otentikasi. Selain itu, KTP-el dilengkapi chip yang berisi data biometrik berupa sidik jari dan iris mata. Hal ini meningkatkan keamanan dan mencegah pemalsuan identitas.[2].

Identitas Kependudukan Digital (IKD) atau Digital ID merupakan inovasi baru dari Direktorat Jenderal Kependudukan dan Pencatatan Sipil (Dukcapil) Kementerian Dalam Negeri dalam rangka meningkatkan kualitas pelayanan kependudukan di Indonesia. IKD hadir sebagai solusi atas berbagai permasalahan dan keluhan masyarakat terkait penerbitan Kartu Tanda Penduduk elektronik (e-KTP) selama ini.

IKD didefinisikan sebagai informasi elektronik yang digunakan untuk merepresentasikan data dan dokumen kependudukan dalam bentuk digital melalui aplikasi di gawai. Data tersebut antara lain berisi identitas dan data pribadi penduduk yang dapat ditunjukkan melalui QR code. Dengan demikian, masyarakat tidak perlu menyimpan KTP fisik, cukup menunjukkan QR code di aplikasi IKD untuk berbagai keperluan administrasi dan kependudukan.[3]

Penerapan IKD didasari Permendagri No. 72 Tahun 2022 tentang Standar dan Spesifikasi Perangkat Keras, Perangkat Lunak, dan Blangko KTP Elektronik serta Penyelenggaraan IKD. Kehadiran IKD diharapkan dapat menjawab tantangan reformasi administrasi kependudukan di era digital saat ini. IKD bakal memudahkan masyarakat dalam mendapatkan hak-hak administrasi seperti mendapatkan dokumen kependudukan, mengakses layanan publik, dan lainnya. Dengan demikian, IKD berpotensi menjadi game changer dalam tata kelola identitas penduduk Indonesia ke depannya.[4]

Dengan demikian, penerapan identitas kependudukan digital memiliki landasan hukum yang jelas dan bertujuan untuk meningkatkan kualitas data kependudukan di Indonesia. Hal ini penting untuk menjamin hak dasar penduduk serta mendukung pelayanan publik yang lebih baik. Identitas Kependudukan Digital (IKD) diharapkan dapat meningkatkan kualitas pelayanan administrasi kependudukan bagi masyarakat. Namun demikian, penerapan sistem baru seperti IKD juga tidak lepas dari berbagai tantangan. Salah satunya adalah respon dan persepsi masyarakat terhadap penerapan IKD yang perlu dipantau dan dievaluasi. Oleh karena itu, penelitian mengenai sentimen masyarakat terhadap IKD perlu dilakukan sebagai evaluasi kebijakan dan umpan balik bagi pemerintah.

Penelitian ini akan melakukan analisis sentimen terhadap opini dan tanggapan masyarakat terkait penerapan IKD di review Google PlayStore. Analisis sentimen merupakan studi komputasional tentang opini, sikap, dan perasaan pada suatu objek tertentu yang diekspresikan dalam teks.[5] Dengan analisis sentimen, dapat diketahui persepsi positif atau negatif masyarakat sehingga dapat menjadi masukan untuk perbaikan dan pengembangan IKD ke depannya.

Oleh karena itu, analisis sentimen terhadap tanggapan dan ulasan masyarakat mengenai IKD dapat dilakukan. Dengan menerapkan teknik text mining dan machine learning, sentimen positif atau negatif masyarakat terhadap penerapan IKD dapat diketahui.[6] Hasil analisis sentimen ini dapat menjadi masukan berharga bagi pemerintah untuk evaluasi dan perbaikan sistem IKD agar lebih inklusif dan sesuai kebutuhan masyarakat Indonesia. Dengan demikian, analisis sentimen diharapkan dapat mendukung keberhasilan implementasi kebijakan IKD di Indonesia.

Analisis sentimen juga dapat diterapkan pada aplikasi Identitas Kependudukan Digital (IKD). Dalam konteks ini, analisis sentimen dapat membantu memahami opini pengguna tentang aplikasi IKD. Dengan menerapkan teknik text mining dan machine learning pada data ulasan pengguna IKD di Google Playstore, sentimen positif atau negatif pengguna terhadap fitur dan kinerja aplikasi IKD dapat dipahami.[7] Hasil analisis sentimen ini nantinya diharapkan dapat menjadi feedback yang berharga bagi pengembang aplikasi IKD dalam upaya peningkatan kualitas dan fitur aplikasi. Dengan demikian, analisis sentimen diharapkan dapat berkontribusi dalam evaluasi dan penyempurnaan aplikasi IKD agar lebih user-friendly dan sesuai dengan kebutuhan masyarakat Indonesia. Pada akhirnya, penerapan analisis sentimen diharapkan turut mendukung keberhasilan implementasi kebijakan Identitas Kependudukan Digital di Indonesia.

Teknik analisis sentimen yang digunakan pada aplikasi Identitas Kependudukan Digital (IKD) dapat melibatkan pengumpulan ulasan atau feedback dari pengguna yang telah menggunakan aplikasi tersebut. Setelah itu, algoritma analisis sentimen akan diterapkan untuk mengklasifikasikan ulasan tersebut sebagai positif, negatif, atau netral[8].

Analisis sentimen dapat membantu pemerintah selaku penyelenggara IKD dalam memahami opini pengguna tentang aplikasi IKD dan bagaimana aplikasi tersebut dapat memenuhi kebutuhan masyarakat. Proses analisis sentimen pada aplikasi IKD melibatkan pengumpulan data ulasan pengguna di Google Playstore, yang kemudian diklasifikasikan menggunakan algoritma analisis sentimen seperti naive bayes, SVM, atau deep learning.

Setelah review aplikasi IKD berhasil dikelompokkan ke dalam sentimen positif dan negatif, topik-topik yang dibahas dalam masing-masing kelompok dapat diidentifikasi. Hal ini memberi gambaran mengenai harapan pengguna terhadap aplikasi IKD dan aspek yang perlu diperbaiki. Dengan demikian, pemerintah dapat meningkatkan kualitas aplikasi IKD sesuai kebutuhan pengguna dan meningkatkan kepuasan masyarakat.

Selain itu, hasil analisis sentimen juga memungkinkan pemerintah membandingkan aplikasi IKD dengan aplikasi sejenis di negara lain yang telah menerapkan identitas digital. Perbandingan ini membantu mengevaluasi kualitas aplikasi IKD dan menentukan pengembangan yang diperlukan agar sesuai standar internasional. Singkatnya, analisis sentimen penting dilakukan untuk mengevaluasi dan menyempurnakan aplikasi IKD di Indonesia.

Dalam penelitian ini, teknik klasifikasi Naive Bayes dan K-Nearest Neighbor akan digunakan untuk melakukan analisis sentimen pada data ulasan pengguna aplikasi Identitas Kependudukan Digital (IKD). Pemilihan kedua metode ini didasarkan pada kelebihan dan karakteristiknya masing-masing dalam mengklasifikasikan data tekstual.[9]

Naive Bayes merupakan classifier probabilistik sederhana yang menerapkan Teorema Bayes dengan asumsi independensi (naive) antar fitur.[10] Metode ini sangat cocok untuk klasifikasi dokumen dan analisis sentimen karena proses training-nya yang cepat dan tidak memerlukan banyak data latih dibandingkan metode lain.[11] Dalam menerapkan Naive Bayes, data ulasan IKD akan melalui proses ekstraksi fitur terlebih dahulu, seperti pemilihan kata, frasa, dan pembobotan tf-idf. Kemudian, peluang prior dan peluang kondisional fitur-fitur tersebut akan dihitung berdasarkan data latih yang telah diberi label kelas sentimen secara manual. Dengan mengalikan peluang-peluang tersebut, data uji dapat diklasifikasikan ke dalam kelas sentimen positif, negatif.[12]

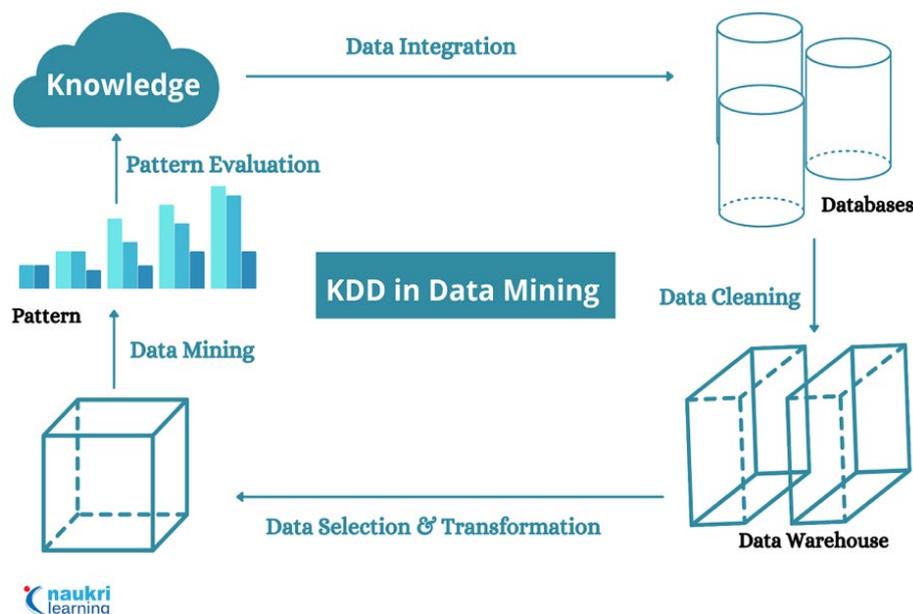
Sementara itu, K-Nearest Neighbor (KNN) merupakan metode klasifikasi berbasis jarak (distance-based). Data uji akan dikategorikan ke dalam kelas yang paling banyak ditemui di antara k tetangga terdekatnya pada data latih. Kelebihan KNN adalah sederhana, mudah diimplementasikan, dan akurasi yang

didapat sebanding dengan jumlah tetangga ( $k$ ) yang dipertimbangkan.[13] Oleh karena itu, nilai  $k$  perlu dituning agar didapatkan model KNN yang optimal untuk klasifikasi sentimen pada ulasan IKD. Dengan menerapkan dan membandingkan kedua metode tersebut, diharapkan analisis sentimen pada ulasan aplikasi IKD dapat dilakukan secara komprehensif dan akurat sehingga memberikan wawasan bermanfaat bagi pengembangan aplikasi ke depannya.

## 2. METODE PENELITIAN

Dalam penelitian ini, metode Knowledge Discovery in Databases (KDD) digunakan sebagai metodologi untuk melakukan analisis sentimen terhadap data ulasan aplikasi IKD. Pada tahap pemilihan data, ulasan pengguna aplikasi IKD di Google Play Store diambil sebagai sumber data yang relevan. Selanjutnya data ulasan tersebut dipreprocess untuk membersihkan noise dan inkonsistensi. Proses ekstraksi fitur dilakukan untuk mentransformasi data tekstual menjadi bentuk numeric yang sesuai input algoritma data mining.

(KDD) Knowledge Discovery in Databases adalah proses mengidentifikasi pola yang sah, baru, dan berpotensi berguna serta dapat dimengerti dari sekumpulan data. Tujuan KDD adalah untuk menemukan pengetahuan berharga dan tersembunyi dari data.[14] Teknik data mining berupa klasifikasi sentimen diimplementasikan untuk menganalisis ulasan dan mengelompokkannya ke dalam sentimen positif, negatif, atau netral. Hasil klasifikasi dievaluasi menggunakan matriks kebingungan dan teknik validasi silang untuk mengukur akurasi. Pola dan pengetahuan yang diperoleh dari hasil klasifikasi sentimen kemudian divisualisasikan agar mudah ditafsirkan. Dengan menerapkan metode KDD yang sistematis, analisis sentimen pada ulasan aplikasi IKD dapat dilakukan secara komprehensif.



**Gambar 1.** Tahapan Metode (KDD) Knowledge Discovery in Databases [15]

### 1. *Understanding the Data Set*

Pada tahap pertama ini, pemahaman mendalam terhadap data ulasan aplikasi IKD yang akan dianalisis perlu dilakukan. Data ulasan pengguna aplikasi IKD diambil dari Google Play Store. Oleh karena itu, penting untuk memahami karakteristik dan struktur data ulasan di platform tersebut. Informasi seperti panjang ulasan, bahasa yang digunakan, skor rating aplikasi, dan lainnya perlu diketahui.

Selanjutnya, tujuan dari analisis sentimen perlu didefinisikan dengan jelas, yaitu untuk mengelompokkan ulasan ke dalam sentimen positif, negatif, atau netral. Batasan dan asumsi terkait kualitas data, skema pengkodean sentimen, serta teknik analisis yang akan digunakan perlu ditetapkan. Dengan memahami data ulasan secara mendalam, analisis sentimen dapat dilakukan secara efektif dan menghasilkan wawasan bermanfaat terkait persepsi pengguna terhadap aplikasi IKD.

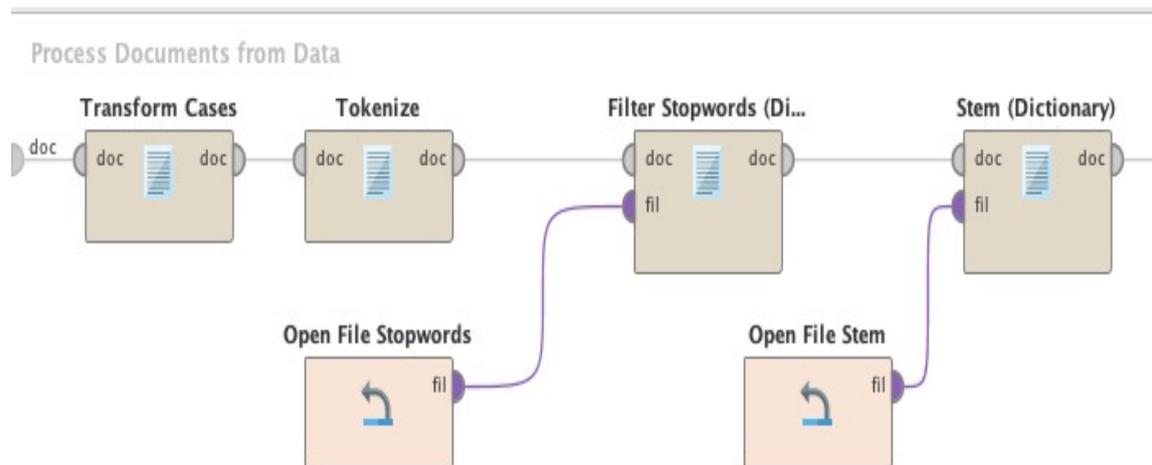
1. *Data Selection*

Setelah tujuan analisis sentimen ditetapkan pada tahap pertama, selanjutnya adalah memilih data ulasan yang relevan dari Google Play Store untuk digunakan dalam penelitian. Tidak semua data ulasan yang tersedia perlu diambil, hanya data ulasan yang memenuhi kriteria tertentu saja yang akan dipilih.

2. *Cleaning and Pre-processing*

Pada tahap ini, data ulasan yang telah terkumpul perlu dipreprocess agar siap digunakan untuk analisis sentimen. Cleaning dilakukan dengan menghapus data yang tidak lengkap, mengandung banyak noise, atau outliers. Misalnya menghapus ulasan kosong atau ulasan yang hanya berisi spam[16].

Selanjutnya dilakukan *text preprocessing* seperti *transform case, case folding, cleaning punctuation & numbers, tokenization, filtering, normalization, stemming, dan stopwords removal*. Tujuannya adalah untuk menstandarisasi data tekstual agar lebih terstruktur dan siap diolah. Data yang tidak relevan seperti tanggal, URL, emoji juga dihilangkan.



Gambar 2. pre-processing, Sumber Penelitian 2023

1. Transform cases

Dalam tahap pra-pemrosesan data, transformasi kasus merupakan prosedur penting yang mengubah semua huruf dalam dokumen tekstual menjadi format seragam, baik huruf kecil semua maupun huruf besar semua. Proses standarisasi ini memiliki tujuan mempermudah analisis data dengan menyederhanakan dan menyeragamkan data sebelum diproses lebih lanjut. Transformasi huruf bertujuan memfasilitasi pemrosesan data yang konsisten dan teliti dalam analisis teks.

Table 1. Hasil Transform Cases

Data sebelum diolah	Data sesudah diolah
Sangat membantu dan mempermudah, masyarakat tidak mengalami kesulitan dalam pengurusan dokumen kependudukan. Tuhan memberkati selalu .	sangat membantu dan mempermudah, masyarakat tidak mengalami kesulitan dalam pengurusan dokumen kependudukan. tuhan memberkati selalu

2. *Tokenize*

Tokenisasi merupakan prosedur analisis tekstual di mana rangkaian karakter dalam sebuah kalimat dipisahkan menjadi satuan-satuan yang lebih kecil yang disebut token. Token dapat berupa kata, frasa, tanda baca, atau karakter individu. Melalui tokenisasi, komputer dapat memproses teks dengan cara yang lebih terstruktur karena teks telah dipecah menjadi komponen-komponen dasar. Dalam tokenisasi, tanda baca seperti titik, koma, dan tanda tanya juga kerap dikategorikan sebagai token tersendiri. Dengan demikian, tokenisasi memungkinkan pemrosesan bahasa alami dan pengolahan teks secara otomatis.

Table 2. Hasil *TOKENIZE*

Data sebelum diolah	Data sesudah diolah
sangat membantu dan mempermudah, masyarakat tidak mengalami kesulitan dalam pengurusan dokumen kependudukan. tuhan memberkati selalu	"sangat" , "membantu" , "dan" , "mempermudah" , " , " , "masyarakat" , "tidak" , "mengalami" , "kesulitan" , "dalam" , "pengurusan" , "dokumen" , "kependudukan" , "." , "tuhan" , "memberkati" , "selalu"

3. *Filter Stopword*

Filter stopwords merupakan teknik pra-pemrosesan data tekstual yang bertujuan mengeliminasi kata-kata yang dianggap tidak memberikan kontribusi signifikan terhadap makna dan tujuan analisis teks. Stopword adalah kata-kata umum seperti "yang", "dan", "dari", dan "pada" yang memiliki nilai informasi rendah. Dengan menghilangkan stopwords dari korpus teks, akurasi dan efisiensi proses analisis teks dapat ditingkatkan. Filter stopwords memungkinkan fokus analisis teks tertuju pada kata-kata konten yang lebih informatif. Sehingga, filter stopwords merupakan langkah pra-pemrosesan penting dalam analisis teks otomatis.

**Table 3.** HASIL Filter Stopword

Data sebelum diolah	Data sesudah diolah
sangat membantu dan mempermudah, masyarakat tidak mengalami kesulitan dalam pengurusan dokumen kependudukan. tuhan memberkati selalu	membantu mudah, masyarakat alami kesulitan urusan dokumen penduduk. tuhan berkati selalu

3. Stemming

Stemming pada analisis sentimen mengacu pada proses mengubah kata-kata dalam teks menjadi bentuk dasar atau kata dasar. Proses ini dilakukan untuk memperoleh pemahaman yang lebih baik tentang makna yang terkandung dalam teks. Dalam konteks analisis sentimen, stemming dapat membantu dalam mengidentifikasi kata kunci yang berkaitan dengan perasaan atau sentimen yang diungkapkan dalam teks. Dengan demikian, stemming dapat membantu dalam memperoleh informasi yang lebih akurat tentang sentimen yang diungkapkan dalam teks.

**Table 4.** Hasil Stemming

Data sebelum diolah	Data sesudah diolah
sangat membantu dan mempermudah, masyarakat tidak mengalami kesulitan dalam pengurusan dokumen kependudukan. tuhan memberkati selalu	sangat bantu mudah, masyarakat tidak alami kesulitan dalam urusan dokumen kependudukan. tuhan berkati selalu

4. *Data Transformation*

Pada tahap transformasi data, data tekstual hasil preprocessing diubah menjadi format numeric yang sesuai sebagai input untuk algoritma data mining. Teknik yang dilakukan di antaranya pembobotan kata (TF-IDF), ekstraksi fitur, dan dimensionality reduction.

5. *Select the Appropriate Data Mining Task*

Pada tahap ini, teknik data mining yang sesuai dengan tujuan analisis sentimen perlu dipilih. Seperti yang telah ditetapkan sebelumnya, tujuan dari penelitian ini adalah mengklasifikasikan ulasan aplikasi IKD ke dalam sentimen positif, negatif, atau netral. Oleh karena itu, teknik data mining yang tepat adalah klasifikasi teks. Algoritma klasifikasi yang dapat digunakan di antaranya Naive Bayes, KNN.

### 3. HASIL DAN PEMBAHASAN

Data ulasan aplikasi Identitas Kependudukan Digital (IKD) dikumpulkan melalui web scraping menggunakan library google play scraper. Proses tersebut berhasil mengumpulkan 2000 ulasan pengguna. Data mentah ini memuat beberapa atribut seperti id, nama pengguna, tanggal, isi ulasan, dan rating aplikasi.

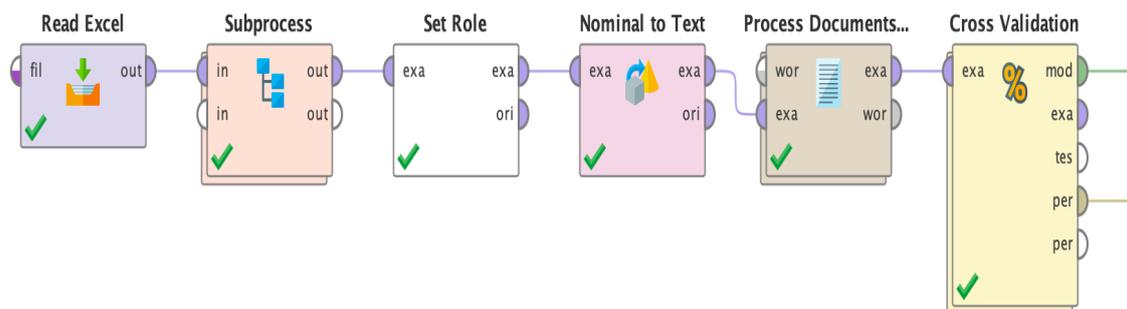
Tahap berikutnya adalah ekstraksi isi ulasan saja sebagai objek analisis sentimen. Ulasan kemudian dilabeli secara manual oleh peneliti ke dalam 2 kelas yaitu positif dan negatif. Didapatkan 1000 ulasan positif dan 1000 ulasan negatif yang akan digunakan. Dengan pra-pemrosesan ini, data siap untuk dibersihkan, ditransformasikan, dan diklasifikasikan untuk mendapatkan sentimen pengguna terhadap aplikasi IKD. Diharapkan hasil analisis sentimen dapat memberi wawasan guna pengembangan aplikasi yang lebih baik lagi di masa mendatang.

**Table 4.** Jumlah Data

Label	Jumlah
<i>Positive</i>	1000
<i>Negative</i>	1000

Pra-pemrosesan teks merupakan langkah fundamental dalam text mining guna menyiapkan data sebelum analisis. Tahapan ini melibatkan serangkaian teknik untuk membersihkan dan menstandarisasi data tekstual. Penelitian ini menerapkan text preprocessing pada data menggunakan perangkat lunak RapidMiner.

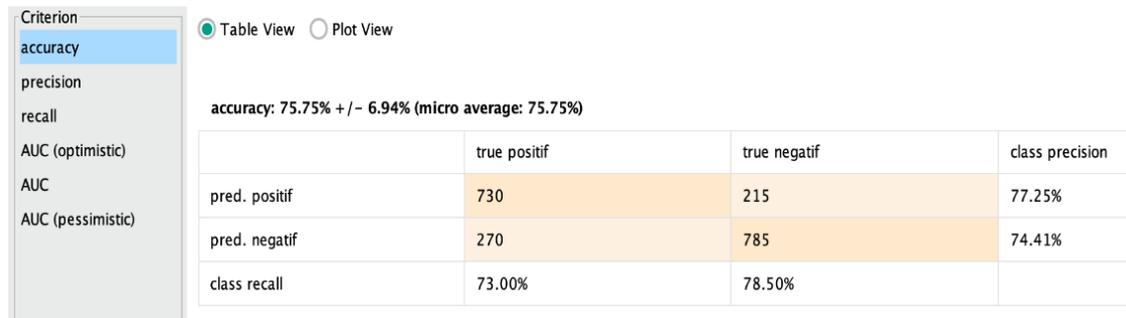
Beberapa prosedur yang dilakukan meliputi penghapusan redundansi, tokenisasi, filtering stopword, stemming, hingga penerapan model spesifik untuk memastikan kualitas data. Tujuannya adalah untuk menghasilkan korpus teks bersih tanpa noise, seragam, dan siap diolah lebih lanjut. Dengan pra-pemrosesan yang komprehensif, kualitas analisis sentimen pada data teks diharapkan dapat optimal dan bermanfaat untuk menggali informasi sentimen pengguna.



**Gambar 3.** Modelling, Sumber Penelitian 2023

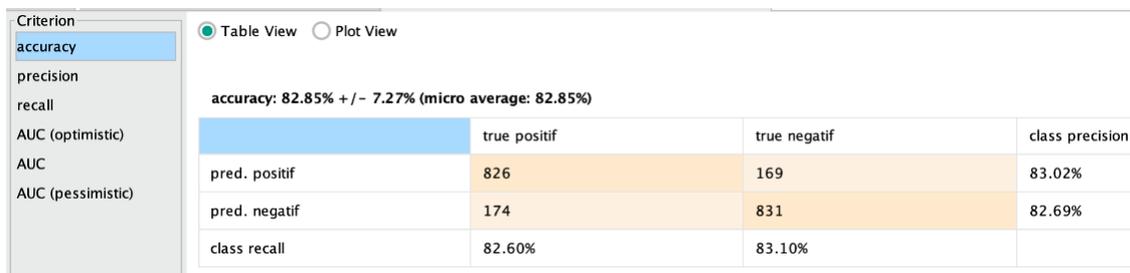
Pemodelan merupakan fase inti dari text mining di mana berbagai algoritma diterapkan untuk mengekstraksi pola dan pengetahuan dari data. Penelitian ini membandingkan dua teknik,

yakni Naive Bayes dan K-Nearest Neighbor (KNN), untuk klasifikasi sentimen. Naive Bayes menggunakan pendekatan probabilistik dengan asumsi independensi prediktor. Sementara KNN menerapkan prinsip kedekatan jarak untuk menentukan kelas data baru berdasarkan tetangga terdekat pada data latih. Kedua algoritma akan diimplementasikan dan dievaluasi performanya untuk mendapatkan model terbaik bagi klasifikasi sentimen. Perbandingan secara komprehensif di antara Naive Bayes dan KNN diharapkan dapat menghasilkan wawasan mengenai teknik yang optimal untuk menganalisis sentimen data tekstual. Pemodelan yang matang penting demi menghasilkan klasifikasi sentimen yang akurat dan bermanfaat dalam penggalan opini pengguna.



Gambar 1. Hasil Akurasi Naive Bayes, Sumber Penelitian 2023

Berdasarkan hasil percobaan, metode Naive Bayes menunjukkan performa yang cukup baik dalam melakukan klasifikasi sentimen pada data ulasan aplikasi Identitas Kependudukan Digital (IKD). Dari matriks kebingungan yang dihasilkan, diketahui bahwa akurasi deteksi sentimen positif mencapai 73% sedangkan akurasi deteksi sentimen negatif mencapai 78,5%. Ini mengindikasikan bahwa secara umum metode Naive Bayes mampu mengklasifikasikan ulasan ke dalam kelas sentimen yang sesuai dengan tingkat akurasi rata-rata sekitar 75%. Meskipun demikian, nilai presisi untuk masing-masing kelas sentimen masih berpotensi untuk ditingkatkan yaitu 77,25% untuk kelas positif dan 74,41% untuk kelas negatif. Dengan kata lain, proporsi klasifikasi yang benar dari keseluruhan hasil prediksi model Naive Bayes masih dapat dikembangkan. Oleh karena itu, penelitian lebih lanjut dengan memodifikasi parameter model dan fitur inputan disarankan agar performa klasifikasi sentimen menggunakan Naive Bayes dapat dimaksimalkan.



Gambar 5. Hasil Akurasi K-NN, Sumber Penelitian 2023

Berdasarkan hasil eksperimen yang dilakukan, metode K-Nearest Neighbor (KNN) menunjukkan performa klasifikasi sentimen yang sangat baik pada data ulasan aplikasi Identitas Kependudukan Digital (IKD). Dari matriks kebingungan yang dihasilkan, diketahui bahwa rata-rata akurasi deteksi sentimen mencapai 82,85%. Angka ini melampaui akurasi rata-rata metode Naive Bayes sebelumnya yang sekitar 75%. Lebih rinci, akurasi deteksi sentimen positif oleh

KNN mencapai 82,6% sedangkan akurasi deteksi sentimen negatif mencapai 83,1%. Nilai presisi untuk kedua kelas sentimen juga cukup tinggi yakni 83,02% untuk positif dan 82,69% untuk negatif. Ini menunjukkan bahwa secara umum KNN mampu memprediksi sentimen ulasan dengan sangat akurat dan presisi yang baik.

#### 4. KESIMPULAN

Penelitian ini memiliki tujuan yang jelas untuk menganalisis sentimen pada data ulasan aplikasi Identitas Kependudukan Digital (IKD) dan membandingkan performa dua metode machine learning, yaitu Naive Bayes dan K-Nearest Neighbor (KNN). Sentimen analisis menjadi sangat penting dalam konteks aplikasi teknologi karena dapat membantu pengembang untuk memahami bagaimana pengguna merespons produk mereka, termasuk aplikasi IKD, dan mengidentifikasi area yang perlu ditingkatkan.

Hasil penelitian menunjukkan bahwa KNN secara konsisten mengungguli Naive Bayes dalam hal akurasi dan presisi. Hal ini mengindikasikan bahwa pendekatan KNN yang berbasis pada jarak antar data latih lebih mampu memahami pola-pola kompleks dalam data ulasan, yang dapat memberikan hasil klasifikasi yang lebih akurat dan konsisten. Penjelasan mengenai keunggulan KNN yang mempertimbangkan kedekatan antar data latih juga menambahkan nilai dan pemahaman lebih mendalam terhadap hasil penelitian. Namun, demikian, penelitian ini tetap memberikan ruang untuk peningkatan lebih lanjut. Salah satunya adalah dengan mengoptimalkan parameter dan hyperparameter dari kedua metode yang digunakan. Selain itu, perlu juga dieksplorasi apakah kombinasi dari berbagai metode seperti SVM dan deep learning dapat meningkatkan performa klasifikasi lebih lanjut.

Keterbatasan penelitian ini juga dapat diatasi dengan mengadopsi variasi data uji yang lebih besar dan komprehensif. Dengan mengumpulkan lebih banyak data ulasan dari berbagai sumber dan kelompok pengguna, hasil analisis sentimen dapat menjadi lebih general dan representatif. Selain itu, pendekatan yang lebih granular terhadap label sentimen juga dapat memberikan wawasan yang lebih detail dan nuansa dalam menganalisis respon pengguna terhadap aplikasi IKD.

Dengan meningkatkan kualitas analisis sentimen pada ulasan aplikasi IKD, pengembang dapat mendapatkan wawasan berharga tentang kekuatan dan kelemahan aplikasi mereka, serta pandangan pengguna secara menyeluruh. Informasi ini dapat digunakan untuk melakukan perbaikan dan peningkatan aplikasi IKD, sehingga dapat lebih responsif terhadap kebutuhan dan harapan masyarakat pengguna. Secara keseluruhan, penelitian ini memberikan kontribusi yang berharga dalam bidang analisis sentimen dan aplikasi teknologi. Namun, seperti kebanyakan penelitian, ada potensi untuk pengembangan lebih lanjut. Dengan melibatkan lebih banyak metode, mengoptimalkan parameter, dan memperluas sampel data, penelitian selanjutnya dapat memperkuat hasil yang dicapai dalam penelitian ini dan membuka jalan untuk pemahaman yang lebih mendalam tentang sentimen pengguna terhadap aplikasi IKD.

#### UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

#### DAFTAR PUSTAKA

- [1] D. C. Tinggogoy and J. H. Kaloly, "PELAYANAN PUBLIK: Antara Hak Masyarakat dan Kewajiban Pemerintah (Studi di Kecamatan Loloda Utara, Kabupaten Halmahera Utara)," *JIPAGS (Journal of Indonesian Public Administration and Governance Studies)*, vol. 5, no. 1, 2021, doi: 10.31506/jipags. v5i1.10333.
- [2] Zefanya Yosua Jocom, Rossy Lambelanova, and Marja Sinurat, "EVALUASI KEBIJAKAN PELAYANAN KARTU TANDA PENDUDUK ELEKTRONIK PADA SUKU DINAS KEPENDUDUKAN DAN PENCATATAN SIPIL KOTA JAKARTA UTARA PROVINSI DKI JAKARTA," *VISIONER: Jurnal Pemerintahan Daerah di Indonesia*, vol. 13, no. 2, 2021, doi: 10.54783/jv. v13i2.437.
- [3] "Mengenal apa itu Identitas Kependudukan Digital-ANTARA News." <https://www.antaranews.com/berita/3392865/mengenal-apa-itu-identitas-kependudukan-digital> (accessed Jul. 24, 2023).
- [4] "Identitas Kependudukan Digital | Disdukcapil." <https://disdukcapil.sidoarjokab.go.id/informasi/identitas-kependudukan-digital> (accessed Jul. 24, 2023).
- [5] K. Anwar, "Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 4, 2022, doi: 10.30865/klik. v2i4.315.
- [6] B. Laurensz and Eko Sedyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 10, no. 2, 2021, doi: 10.22146/jnteti. v10i2.1421.
- [7] S. R. Wardhana and D. Purwitasari, "Klasifikasi Multi Class Pada Analisis Sentimen Opini Pengguna Aplikasi Mobile Untuk Evaluasi Faktor Usability," *INTEGER: Journal of Information Technology*, vol. 4, no. 1, 2019, doi: 10.31284/j.integer.2019. v4i1.474.
- [8] D. I. Af'idah, D. Dairoh, S. F. Handayani, R. W. Pratiwi, and S. I. Sari, "Sentimen Ulasan Destinasi Wisata Pulau Bali Menggunakan Bidirectional Long Short Term Memory," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, 2022, doi: 10.30812/matrik. v21i3.1402.
- [9] N. B. Putri and A. W. Wijayanto, "Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing," *Komputika: Jurnal Sistem Komputer*, vol. 11, no. 1, 2022, doi: 10.34010/komputika. v11i1.4350.
- [10] H. Hafizah, T. Tugiono, and A. Azlan, "Sistem Pakar Untuk Pendiagnosaan Karies Gigi Menggunakan Teorema Bayes," *J-SISKO TECH (Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD)*, vol. 4, no. 1, 2021, doi: 10.53513/jsk. v4i1.2625.
- [11] A. P. Widyassari and P. E. Suryani, "Komparasi Metode Naïve Bayes dan SAW untuk Pemilihan Penerimaan Insentif Karyawan," *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, vol. 3, no. 02, 2021, doi: 10.46772/intech. v3i02.555.
- [12] I. B. N. W. Manuaba, G. R. Dantes, and G. Indrawan, "Analisis Sentimen Data Provider Layanan Internet Pada Twitter Menggunakan Support Vector Machine Dengan Penambahan Algoritma Levenshtein Distance," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 5, no. 2, 2022, doi: 10.47970/siskom-kb. v5i2.261.
- [13] N. M. Putri, "KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELLITUS," *EVOLUSI: Jurnal Sains dan Manajemen*, vol. 10, no. 1, 2022, doi: 10.31294/evolusi. v10i1.12514.
- [14] M. A. Sembiring, R. A. T. Agus, and M. F. L. Sibuea, "ANALISIS KEPUASAN PELANGGAN MENGGUNAKAN METODE ROUGH SET," *JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, vol. 4, no. 2, 2021, doi: 10.54314/jssr. v4i2.647.

- [15] “Knowledge Discovery in Databases (KDD) in Data Mining.” <https://www.shiksha.com/online-courses/articles/kdd-in-data-mining/> (accessed Jul. 25, 2023).
- [16] B. Widodo, H. A. Armanto, and E. Setyati, “Deteksi Pemakaian Helm Proyek Dengan Metode Convolutional Neural Network,” *Journal of Intelligent System and Computation*, vol. 3, no. 1, 2021, doi: 10.52985/insyst.v3i1.157.