YAYASAN AKRAB PEKANBARU
Akrab Juara : Jurnal Ilmu-ilmu Sosial
Volume 11 Nomor 2 Edisi Mei 2026 (360-367)

SENTIMENT ANALYSIS OF DIGITAL POPULATION IDENTITY (IKD) APPLICATION USING THE NAÏVE BAYES ALGORITHM

Eka Putra Pratama, Ade Setiawan, Rifky Permana, Sandra Dewi Saraswati, Tri Hartati
Universitas Bina Sarana Informatika
(Naskah diterima: 1 Maret 2026, disetujui: 31 Maret 2026)

Abstract

This study aims to analyze user sentiment towards the digital identity application launched by the government. The methodology employed involves collecting user reviews and comments from various social media platforms and the official application. The data is then analyzed using Natural Language Processing (NLP) techniques to identify and categorize sentiments as positive, negative, or neutral. The Naive Bayes method will be implemented using a dataset containing user reviews about a digital identity application. The analysis results indicate that the majority of user sentiments towards the application are positive, with many users appreciating the ease of access and features that simplify administrative processes. However, there are also a number of negative sentiments indicating technical issues and concerns regarding data privacy. Neutral sentiments generally relate to suggestions for improvements and enhancements to the application. The findings of this study are expected to provide constructive feedback for application developers to improve the quality and security of the digital identity application, as well as to enhance user satisfaction in the future. The study uses Google Colab, a cloud-based service provided by Google that allows users to write and execute Python code directly from the browser. The researcher used the scraping method to collect review data from IKD applications on the Play Store.

Keywords: *Sentiment Analysis of Digital Population Identity (IKD) Application users*

Abstract

Studi ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi identitas digital yang diluncurkan oleh pemerintah. Metodologi yang digunakan melibatkan pengumpulan ulasan dan komentar pengguna dari berbagai platform media sosial dan aplikasi resmi. Data kemudian dianalisis menggunakan teknik Pemrosesan Bahasa Alami (NLP) untuk mengidentifikasi dan mengkategorikan sentimen sebagai positif, negatif, atau netral. Metode Naive Bayes akan diimplementasikan menggunakan dataset yang berisi ulasan pengguna tentang aplikasi identitas digital. Hasil analisis menunjukkan bahwa mayoritas sentimen pengguna terhadap aplikasi tersebut positif, dengan banyak pengguna menghargai kemudahan akses dan fitur yang menyederhanakan proses administrasi. Namun, ada juga sejumlah sentimen negatif yang menunjukkan masalah teknis dan kekhawatiran terkait privasi data. Sentimen netral umumnya berkaitan dengan saran untuk perbaikan dan peningkatan aplikasi. Temuan studi ini diharapkan dapat memberikan umpan balik yang konstruktif bagi pengembang aplikasi untuk meningkatkan kualitas dan keamanan aplikasi identitas digital, serta meningkatkan kepuasan pengguna di masa mendatang. Studi ini menggunakan Google Colab, layanan berbasis cloud yang disediakan oleh Google yang memungkinkan pengguna untuk menulis dan menjalankan kode Python langsung dari

browser. Peneliti menggunakan metode scraping untuk mengumpulkan data ulasan dari aplikasi IKD di Play Store.
Kata kunci: Analisis Sentimen Pengguna Aplikasi Identitas Populasi Digital (IKD)

I. INTRODUCTION

In today's digital era, the government is leveraging technology to improve the efficiency and quality of services provided to the public. One example is the use of digital population identity applications, which aim to replace physical identity documents such as the National Identity Card (KTP). "Digital Population Identity (IKD) is one innovation developed to support the digitization of population documents" (Wahyuningsih, 2023).

These applications allow the public to access and update population data online, with the aim of simplifying administration and data management processes. However, the use of digital population identity applications also involves various aspects that need to be evaluated, including user sentiment towards the application.

In writing this thesis, the author is interested in analyzing the IKD application. The aim is to determine sentiment towards the application by extracting reviews from the Play Store. The results can then be utilized by developers.

"Sentiment analysis can be used to assess a person's opinions and emotions about certain things" (Cendana, 2020).

"Sentiment analysis is a subset of text mining that aims to categorize text documents based on opinions, thus producing sentiment information that can be either positive or negative" (Ardianto et al., 2020).

By analyzing user sentiment towards digital identity applications, we can obtain data on user satisfaction, potential issues, and features that need improvement.

"In this study, text was classified using the naive Bayes method, and the query extension ranking method was used to reduce the number of features used in the classification process. The results showed an accuracy of 86.6%" (Fanissa, 2020).

"This method is considered suitable for use in sentiment analysis because it aims to categorize user opinions or reviews into positive, negative, or neutral" (Arvyantomo, 2023).

It is hoped that the government and application developers can use the results of this sentiment analysis to improve the quality and user satisfaction and address potential issues. The Naive Bayes method will be implemented using a dataset containing user reviews of digital identity applications. investment platform and help companies improve services, then

researchers want to conduct a study entitled: Sentiment Analysis of the Ajaib Application to Determine Satisfaction Based on Opinions Using the BERT Algorithm.

II. RESEARCH METHODS

The research stages involved in sentiment analysis are as follows:

Data Collection: Reviewing digital identity applications and collecting user reviews from various sources, such as the Google Play Store, App Store, and application websites, is the method of data collection.

Data Preprocessing: Preprocessing is the first and most crucial stage in the text mining process. "Preprocessing is carried out to avoid inconsistent or incomplete data" (Muzaki, 2021).

The collected data is then processed to remove noise and convert the text into a format that can be processed by the Naive Bayes Classifier algorithm. Using the Naive Bayes Classifier Algorithm: This algorithm is used to predict user reactions to digital identity applications. The training and testing stages are included in this process.

"Due to its ease of use and fast processing, Naive Bayes is one of the most widely used algorithms for data mining purposes. It is easy to use, has a simple structure, and is highly effective" (Khotimah, 2022).

Results Evaluation: The sentiment analysis results were examined using evaluation methods such as accuracy, precision, recall, and F-measure to determine how well the Naive Bayes Classifier algorithm analyzed user sentiment.

Results Analysis: Then, it was studied how digital population identity applications can improve user satisfaction through sentiment analysis. "In sentiment analysis, the dataset domain is defined, preprocessing is performed, promoted options are labeled, classified, and evaluated" (Suryani, 2019).

III. RESEARCH RESULTS

1. Research Results

In this discussion, the researcher will explain the results of data processing using the naive Bayes method. The following are several steps in data processing:

Source: Research (2024)

Figure IV.1 Illustration of data collection

2. Data Scraping

In this research phase, the researcher will extract IKD app review data from the Google Play Store. This will be done using the google_play_scraper library, extracting review data from the IKD app using the code gov.dukcapil.mobile_id, and using the Sort.NEWEST code to retrieve the most recent comment column data. This research obtained 1,000 data points for processing in the next phase. The results obtained from the data scraping method are as follows:

Figure IV.2 Google Scrapper script code and package. Source: Google Collab data processing.

The following are the results of the data scraping code above:

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
0	c1c4b362-885e-4939-84dd-ab9f8fc918eb	eko Jailani	https://play-lh.googleusercontent.com/a-/ALV-U...	Sekarang tidak bisa di buka,setalu kesalahan k...	5	165	1.2.2	2024-07-04 05:30:42	None	None	1.2.2
1	686e43c1-3d38-46ba-974a-8c4b4946e6a0	Septi triastuti	https://play-lh.googleusercontent.com/a-/ALV-U...	Aplikasi nya "terjadi kesalahan koneksi" ,suda...	1	441	1.2.2	2024-05-02 02:41:54	None	None	1.2.2
2	fd30e08d-3859-4c82-8a91-f8ba9c846256	MUHAMMAD ZAKI ROBBI AKHSIN	https://play-lh.googleusercontent.com/a-/ALV-U...	Aplikasi tidak responsif, berjalan lambat, ms...	2	606	1.2.2	2024-05-15 08:20:22	None	None	1.2.2
3	4abb1bdf-1bf9-4a8e-848f-481ffe926e62	Khairil Aryan TV	https://play-lh.googleusercontent.com/a-/ALV-U...	Apasih gunanya E-KTP tapi gak bisa cetak KTP-n...	1	597	1.2.2	2024-05-01 01:41:45	None	None	1.2.2
4	e389852f-bdf6-4dc0-bc08-a7520fdfaf8c	Abdul Mutholib	https://play-lh.googleusercontent.com/a/ACg8oc...	Aku kasih bintang 3 karna mau masuk aja harus ...	3	94	1.2.2	2024-05-13 06:41:13	None	None	1.2.2

Figure IV.3 Data Scraping Results

Source: Google Collab data processing

The extraction process begins by retrieving reviews from the Play Store using the application code "gov.dukcapil.mobile_id." This process selects the 1,000 most recent reviews (NEWEST) and exports them as a CSV file.

3. Data Repair

Data repair is the process of repairing damaged or incomplete data so that it can be used effectively in analysis. This process is crucial in data management, as low-quality data can lead to inaccurate analysis.

The following is the coding for data repair:

```
data = data[['content','score']]
data.head()
```

Figure IV.4 Data Preparation

Source: Google Colab data processing

After performing the coding above, here are the results:

	content	score
0	Sekarang tidak bisa di buka,selalu kesalahan k...	5
1	Aplikasi nya "terjadi kesalahan koneksi" ,suda...	1
2	Aplikasi tidak responsif, berjalan lambat, res...	2
3	Apasih gunanya E-KTP tapi gak bisa cetak KTP-n...	1
4	Aku kasih bintang 3 karna mau masuk aja harus ...	3

Figure IV.5 Results of Data Repair

Source: Google Collab data processing

4. Data Visualization

Next, the "WordList to Data" operator is connected to the "Process Documents from Data" operator. This is used to calculate the weighted values and the number of words appearing in the preprocessed dataset.

Figure IV.6 Wordcloud

Source: Google Collab data processing

The following is the wordcloud coding:

```
# Menggabungkan semua ulasan menjadi satu teks
all_text = ' '.join(data['clean_text'])

# Membuat Word Cloud
wordcloud = WordCloud(width=1000, height=500, max_font_size=150, random_state=42).generate(all_text)

# Menampilkan Word
plt.figure(figsize=(10, 6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

Figure IV.7 Script code from Wordcloud

Source: Google Collab data processing

5. Sentiment Analysis Results

The sentiment analysis results can provide a general overview of how the application is received by the public, whether there are more positive or negative reviews, and specific aspects frequently mentioned in reviews.

The following are sentiment labels from user reviews of the IKD application:

	ulasan	sentiment_label
0	Sekarang tidak bisa di buka,selalu kesalahan k...	netral
1	Aplikasi nya "terjadi kesalahan koneksi" ,suda...	negatif
2	Aplikasi tidak responsif, berjalan lambat, res...	netral
3	Apasih gunanya E-KTP tapi gak bisa cetak KTP-n...	netral
4	Aku kasih bintang 3 karna mau masuk aja harus ...	netral
5	Sangat,,,, sangat,,, sangat jelek & buruk,,, u...	netral
6	Sebenarnya ini untuk mempermudah apa gimana si...	netral
7	Aplikasi apa ini, tidak membantu sama sekali,,...	netral
8	tujuan dari aplikasi ini mungkin sudah baik.....	netral
9	Yang seharusnya memudahkan, jadinya bikin tamb...	netral

Figure IV.8 Sentiment Label of IKD user reviews Source: Google Collab data processing

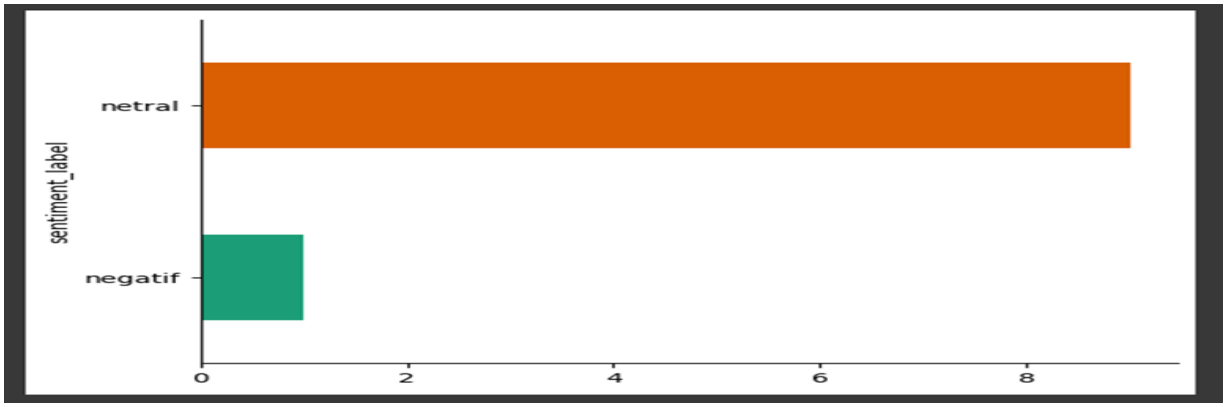

Figure IV.9 Graph of Label Sentiment Results

Source: Google Collab Data Processing

IV. CONCLUSION

1. In the study, entitled "Sentiment Analysis of the Ajaib App to Determine Satisfaction Based on Opinions Using the BERT Algorithm," 1,393 data sets were used, with 696 positive reviews and 697 negative reviews. This study concluded that Ajaib app users had more negative reviews. The results showed an accuracy of 85%, an F1 Score of 85%, a recall of 85%, and a precision of 87%.
2. This study demonstrated a relationship between information and user satisfaction, enabling the company to understand user needs and evaluate them in the future.
3. This study found a significant number of negative reviews, which could impact user trust in the Ajaib app. Therefore, the Ajaib app must improve its system in the future to garner more positive reviews and increase user trust.

REFERENCE

Ardianto, Rian, Tri Rivanie, Yuris Alkhalifi, Fitra Septia Nugraha, and Windu Gata. 2020. "Sentiment Analysis on E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine." Jurnal Ilmu Komputer Dan Informasi 13, no. 2: 109–22. https://doi.org/10.21609/jiki.v13i2.885.

Arvyantomo, M, and N Ratama. 2023. "Analisis Sentimen Masyarakat Indonesia Terhadap Invasi Russia Di Ukraina Menggunakan Metode Naive Bayes Pada Media Sosial Facebook."LOGIC:JurnalIlmuKomputer…1,no.4:705–17. https://www.journal.mediapublikasi.id/index.php/logic/article/view/2142.

Atmadja, Boby Rizki. 2022. "Analisis Sentimen Bahasa Indonesia Pada Tempat Wisata Di Kabupaten Sukabumi Dengan Naive Bayes Classifier." Elkom : JurnalElektronikaDanKomputer15,no.2:371–82. https://doi.org/10.51903/elkom.v15i2.872.

Cendana, Maya, and Silvester Dian Handy Permana. 2019. "Pra-Pemrosesan Teks Pada Grup Whatsapp Untuk Pemodelan Topik." Junal Mantik Penusa 3, no. 3: 107–16.

Dwi Antonio, V., Efendi, S., & Mawengkang, H. (2022). Sentiment analysis for covid-19 in Indonesia on Twitter with TF-IDF featured extraction and stochastic gradient descent. Int. J. Nonlinear Anal. Appl, 13(1), 2008– 6822. https://doi.org/10.22075/IJNAA.2021.5735

Fauziningrum, E., & Suryaningsih, E. I. (2021). Evaluasi Dan Prediksi Penguasaan Bahasa Inggris Maritim

Fanissa, Shima, M Ali Fauzi, and Sigit Adinugroho. 2018. "Analisis Sentimen Pariwisata Di Kota Malang Menggunakan Metode Naive Bayes Dan Seleksi Fitur Query Expansion Ranking." Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer 2, no. 8: 2766–70. http://j-ptiik.ub.ac.id.

Hakimi, Fajar Darwis Dzikril. 2018. "Sistem Analisis Sentimen Publik Tentang Opini Pemilihan Kepala Daerah Jawa Timur 2018 Pada Dokumen Twitter Menggunakan Naive Bayes Classifier." Skripsi Universitas Islam Negeri Sunan Ampel Surabaya.

Halim, A., & Andri Safuwan. (2023). Analisis Sentimen Opini Warganet Twitter Terhadap Tes Screening Genose Pendeteksi Virus Covid-19 Menggunakan Metode Naïve Bayes Berbasis Particle Swarm Optimization. Jurnal InformatikaTeknologiDanSains,5(1),170–178. https://doi.org/10.51401/jinteks.v5i1.2229

Jaya Hidayat, T. H., Ruldeviyani, Y., Aditama, A. R., Madya, G. R., Nugraha, A. W., & Adisaputra, M. W. (2021). Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier. Procedia Computer Science, 197(2022), 660–667. https://doi.org/10.1016/j.procs.2021.12.187

Khotimah, Aristin Chusnul, and Ema Utami. 2022. "Perbandingan Algoritma Naïve Bayes Classifier, K-Nearest Neighbor Dan Support Vector Machine Dalam Klasifikasi Karakter Individu Pada Akun Twitter." Jurnal Teknik Informatika (Jutif)3,no.3:673–80. http://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/254.

Muzaki, Akhmad, and Arita Witanti. 2021. "Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive BayesClassifier."JurnalTeknikInformatika(Jutif)2,no.2:101–7. https://doi.org/10.20884/1.jutif.2021.2.2.51.

Prasetyo, Tri, Hadi Zakaria, and Pandu Wiliantoro. 2022. "Analisis Layanan Pelanggan PT PLN Berdasarkan Media Sosial Twitter Dengan Menggunakan Metode Naïve Bayes Classifier." OKTAL : Jurnal Ilmu Komputer Dan Sains 1, no. 6: 573–82. https://journal.mediapublikasi.id/index.php/oktal.

Suryani, Putu Sri Merta, Linawati Linawati, and Komang Oka Saputra. 2019. "Penggunaan Metode Naïve Bayes Classifier Pada Analisis Sentimen Facebook Berbahasa Indonesia."MajalahIlmiahTeknologiElektro18,no.1:145. https://doi.org/10.24843/mite.2019.v18i01.p22.

Setya Ananto, F., & Hasan, F. N. (2023). Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store. Jurnal ICT : Information Communication&Technology,23(1),75–80. https://ejournal.ikmi.ac.id/index.php/jict-ikmi

Sulahmi, Wardani, S., & Fairuzabadi, M. (2023). Web Scraping Untuk Pengambilan Data Transaksi Penjualan Pada Markeplace Toko Online Jogjabatik. JIIFKOM (Jurnal Ilmiah Informatika Dan Komputer), 2(2), 14–17. https://doi.org/10.51901/jiifkom.v2i2.254

Villavicencio, C., Macrohon, J. J., Inbaraj, X. A., Jeng, J.-H., & Hsieh, J.-G. (2021). Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines UsingNaïveBayes.Information,12(5),204. https://doi.org/10.3390/info12050204

Wahyuningsih, N. 2023. "Perbandingan Metode Klasifikasi Dalam Analisis Sentimen Masyarakat Terhadap Identitas Kependudukan Digital (IKD)" 8, no. 4: 1218–27. https://repository.uksw.edu/handle/123456789/30436%0Ahttps://repository.uksw.edu/bitstream/123456789/30436/8/T1_672019296_Daftar Pustaka.pdf.

Wiranti, N. E., & Frinaldi, A. (2023). Meningkatkan Efisiensi Pelayanan Publik dengan Teknologi di Era Digital. Menggunakan Metode Decision Tree Dan Confusion Matrix (Studi Kasus Di UniversitasMaritimAmni).ProsidingKemaritiman2021. http://repository.unimar- amni.ac.id/id/eprint/3837

JIM:JurnalIlmiahMahasiswaPendidikanSejarah,8(2),748–754. https://doi.org/10.24815/jimps.v8i2.24833

Yin, J. Y. B., Saad, N. H. M., & Yaacob, Z. (2022). Exploring Sentiment Analysis on E-Commerce Business: Lazada and Shopee. TEM Journal, 11(4), 1508–1519. https://doi.org/10.18421/TEM114-11

Yousef, M., & ALali, A. (2022). Analysis and Evaluation of Two Feature Selection Algorithms in Improving the Performance of the Sentiment Analysis Model of Arabic Tweets. International Journal of Advanced Computer Science and Applications,13(6),705–711. https://doi.org/10.14569/IJACSA.2022.0130683