

Analisis Sentimen Ulasan Aplikasi Identitas Kependudukan Digital Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor

Bagus Setiawan*, Kiki Ahmad Baihaqi, Euis Nurlaelasari, Hanny Hikmayanti Handayani

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Buana Perjuangan Karawang, Karawang, Indonesia
Email: ^{1,*}if20.bagussetiawan@mhs.ubpkarawang.ac.id, ²kikiahmad@ubpkarawang.ac.id, ³euis.nurlaelasari@ubpkarawang.ac.id, ⁴hanny.hikmayanti@ubpkarawang.ac.id

Email Penulis Korespondensi: if20.bagussetiawan@mhs.ubpkarawang.ac.id

Submitted: 21/06/2024; Accepted: 30/06/2024; Published: 30/06/2024

Abstrak—Pemerintah telah mengeluarkan inovasi terbaru Pengumpulan data dalam ranah data kependudukan yang mengandalkan teknologi digital melalui aplikasi seluler dengan menggunakan foto atau kode QR yang bertujuan untuk mengurangi penggunaan cetakan fisik kartu tanda penduduk serta ketersediaan blanko KTP dengan tujuan untuk menyederhanakan proses administrasi dan dokumen kependudukan tidak lagi memerlukan pencetakan atau penyimpanan dalam format fisik seperti berkas KTP. Dalam mengimplementasikan aplikasi identitas kependudukan beberapa masyarakat merasa cemas karena keterbatasan akses internet, kurangnya pengetahuan tentang aplikasi, serta kekhawatiran akan keamanan dan privasi data identitas dalam format digital. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan aplikasi identitas kependudukan digital dengan membandingkan algoritma logistic regression dan k-nearest neighbor. Dataset yang di ambil menggunakan library google play scraper pada python yang mendapatkan 1700 data mentah yang di ambil pada tanggal 12-februari sampai 26 Maret 2024 lalu di pre-processing dan mendapatkan 1108 data bersih. Hasil penelitian ini mendapatkan perbandingan antara algoritma logistic regression dan k-nearest neighbor bahwa algoritma k-nearest neighbor lebih baik di bandingkan dengan algoritma logistic regression dengan hasil akurasi sebesar 80.43% selisih 3.60% di bandingkan k-nearest neighbor. Dapat di simpulkan bahwa aplikasi identitas kependudukan digital masih tergolong kurang baik dalam penggunaanya karena memiliki sentimen negatif sebanyak 73,9% dan dapat dilihat pada penelitian ini hasil perbandingan algoritma k-nearest neighbor membuktikan bahwa performanya lebih baik di bandingkan logistic regression.

Kata Kunci: IKD; Sentimen; Ulasan; Kependudukan; Logistic Regression; K-Nearest Neighbor

Abstract—The government has launched the latest innovation in data collection in the realm of population data which relies on digital technology through mobile applications using photos or QR codes which aims to reduce the use of physical prints of identity cards and the availability of blank KTPs with the aim of simplifying the administrative process and no longer requiring population documents. printing or saving in physical format such as an KTP file. In implementing the population identity application, some people feel anxious due to limited internet access, lack of knowledge about the application, as well as concerns about the security and privacy of identity data in digital format. This research aims to conduct sentiment analysis on reviews of digital population identity applications by comparing logistic regression and k-nearest neighbor algorithms. The dataset was taken using the Google Play Scraper library in Python which got 1700 raw data taken from 12-February to 26 March 2024 and then pre-processed and got 1108 clean data. The results of this research show that the comparison between the logistic regression algorithm and k-nearest neighbor algorithm shows that the k-nearest neighbor algorithm is better than the logistic regression algorithm with an accuracy result of 80.43%, a difference of 3.60% compared to k-nearest neighbor. So it can be concluded that the digital population identity application is still considered poor in its use because it has a negative sentiment of 73.9% and it can be seen in this research that the comparison results of the k-nearest neighbor algorithm prove that its performance is better than logistic regression.

Keywords: IKD; Sentiment; Review; Population; Logistic Regression; K-Nearest Neighbor

1. PENDAHULUAN

Kemajuan teknologi informasi telah mengubah kebiasaan manusia dalam kehidupannya. Berkembangnya teknologi informasi dan komunikasi telah meningkatkan efektivitas dan efisiensi tugas yang biasanya dilakukan oleh manusia [1]. Selain itu, teknologi informasi kini menjadi bagian dari aspek sosial budaya, pertahanan, ekonomi, dan masyarakat. Teknologi informasi merupakan bagian integral dari pemerintahan. Karena itu, pemerintah terus berinovasi [2] Aplikasi Identitas Kependudukan Digital merupakan salah satu inovasi yang dibuat untuk mendapatkan dokumen kartu tanda penduduk digital [3].

Pembuatan data kependudukan digital dapat diakses melalui aplikasi seluler dengan gambar atau QR Code dikenal dengan Identitas Kependudukan Digital (IKD). Pemerintah ingin meningkatkan efisiensi administrasi dengan mengurangi cetakan fisik Penggunaan blanko KTP Elektronik dan Kartu Tanda Penduduk (KTP) [4] [5]. Pemerintah berharap dengan adanya aplikasi IKD untuk keperluan administrasi, dokumen kependudukan tidak harus disimpan dalam bentuk berkas fisik KTP. Penggunaan blanko KTP elektronik juga dapat menghemat anggaran pemerintah [6].

Kualitas dan efektivitas pengelolaan data populasi diperkirakan akan meningkat dengan perkembangan baru ini, tetapi ada diskusi yang sedang berlangsung mengenai cara untuk menggunakan dan menjaga aplikasi IKD. Beberapa orang, terutama yang berada di daerah terpencil, mengkhawatirkan keterbatasan akses internet dan minimnya interaksi mengenai aplikasi ini. Selain itu, ada kekhawatiran tentang keamanan data identitas digital pada ulasan IKD di play store di mana orang dapat berbagi pendapat dan komentar [7].

Analisis sentimen, juga disebut opini mining atau kecerdasan kreatif emosional, berguna untuk memproses bahasa alami, analisis teks, dan komputasi linguistik. Ini memungkinkan untuk mengidentifikasi, mengekstrak,

menghitung, dan mempelajari data secara sistematis, dan menghasilkan skor f1, akurasi, presisi, dan recall yang dihasilkan oleh algoritma klasifikasi sentimen[8]. Recall menunjukkan hasil presentasi data yang terklasifikasikan dengan benar, Akurasi mengindikasikan tingkat ketepatan model dalam melakukan klasifikasi yang benar, sementara presisi mengukur seberapa tepat data yang diminta sesuai dengan hasil prediksi yang diberikan oleh model [9].

Pada penelitian terdahulu yang telah di lakukan oleh Rita, dkk, ada analisis sentimen terhadap identitas kependudukan digital terkait dengan penggunaan algoritma support vector machine mendapatkan nilai yang baik dengan angka akurasi 77% dan nilai presisi di angka 92% [10]. Penelitian sebelumnya yang dilakukan oleh Anshari, dkk, mengenai komprasi payment digital untuk mengetahui sentimen dari ulasan di Google Play Store, kami menerapkan pendekatan Support Vector Machine (SVM), penelitian ini menghasilkan hasil aplikasi dana yang mendapat nilai tertinggi dengan nilai akurasi 92.0%, presisi 92.0% [11]. Raja, dkk, juga melakukan penelitian tentang analisis pendapat pengguna Twitter mengenai kenaikan harga bahan bakar minyak (BBM) melalui penggunaan teknik logistic regression dan mendapatkan nilai yang memiliki tingkat akurasi yang baik sebesar 77.0%, nilai presisi 95.0% dan menyimpulkan bahwa mayoritas orang mendukung kenaikan harga BBM [12]. Lalu penelitian yang dilakukan oleh Tiara, dkk, menggunakan teknik K-NN untuk mengkategorikan persepsi publik mengenai kenaikan biaya BBM di Twitter. Pada perbandingan data pelatihan dan pengujian 90:10, tingkat akurasi sebesar 83.3% berhasil dicapai baik dengan K=13 maupun K=15 [13]. Dan penelitian yang di lakukan oleh Audenza, dkk, mengenai comparasion of logistic regression, multinomialnb, svm, and k-nn method on sentiment analysis of gojek app reviews on the google playstore. Diperoleh bahwa metode logistic regression ideal untuk membedakan ulasan positif dan negatif berdasarkan data ulasan aplikasi Gojek di google play store dengan nilai akurasi sebesar 82.45%, dan presisi sebesar 82.45% [14].

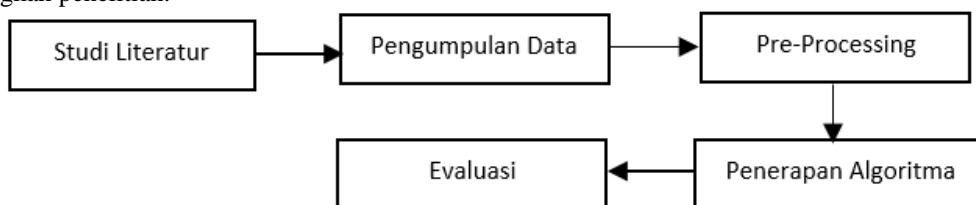
Analisis sentimen ulasan identitas kependudukan digital telah di lakukan sebelumnya menggunakan algoritma non-linear yaitu support vector machine dengan hasil negatif pada ulasan yang menunjukkan ketidakpuasan masyarakat terhadap identitas kependudukan digital dengan persentase 8.76% positif, 12.97 netral, dan 78.27% negatif, oleh karena itu penelitian ini akan mencari hasil dari ulasan identitas kependudukan digital dan perbandingan nilai akurasi dengan algoritma logistic dan k-nearest neighbor karena kedua algoritma tersebut bersifat linear, maka dari itu kedua algoritma tersebut baik untuk dilakukan perbandingan.

Berdasarkan sebagaimana telah dijelaskan, tujuan penelitian ini adalah untuk mengidentifikasi jumlah sentimen positif dan negatif yang ditemukan dalam ulasan aplikasi IKD, serta kata apa yang paling sering ditemukan dalam ulasan aplikasi. Serta membandingkan algoritma Logistic Regression dan juga K-Nearest Neighbor dalam klasifikasi sentimen ulasan pengguna aplikasi Identitas Kependudukan Digital dan juga dapat menjadi acuan, sumber informasi, dan bahan referensi bagi peneliti di masa mendatang dalam proses klasifikasi sentimen.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian adalah serangkaian metode untuk melakukan penelitian. Langkah-langkah penelitian dilakukan guna memperlancar proses penelitian sehingga hasil penelitian dapat tersusun dengan baik. Gambar 1 menunjukkan langkah-langkah penelitian.



Gambar 1. Tahapan penelitian

Pada gambar 1 dapat dilihat Tahap penelitian dimulai dari studi literatur. Kedua, tahap pengumpulan data ulasan pengguna aplikasi. Ketiga, dataset yang telah didapat di preprocessing untuk di lakukan penerapan algoritma. Keempat, implementasi algoritma menggunakan Logistic regression dan K-nearest neighbor. Tahap terakhir tahapan penelitian evaluasi menggunakan confusion matrix.

2.1.1 Studi literatur

Tahap penelitian dimulai dengan melakukan studi literatur yaitu mempelajari dan memahami jurnal yang berkaitan dengan algoritma logistic regression dan K-nearest neighbor dan aplikasi indentitas kependudukan digital yang bersumber dari internet dan buku.

2.1.2 Pengumpulan Data

Pengumpulan data ulasan pengguna aplikasi Identitas Kependudukan Digital dengan library python Google_Play_Scrapper digunakan. Data yang di kumpulkan adalah ulasan terbaru pada tanggal 12 Februari 2024 sampai 26 Maret 2024. Dan mendapatkan data yang berjumlah 1700 data dengan informasi username pengguna, text, dan score ulasan pengguna

2.1.3 Pre-Processing

Tahap ini dalam proses menyiapkan data mentah untuk digunakan atau dianalisis dalam model pembelajaran mesin dilakukan. Tahap ini sangat penting karena kualitas data dan karakteristik data dapat sangat memengaruhi kinerja model. Untuk memilih data yang akan diproses dalam dokumen, tahapan pemrosesan dilakukan untuk mengolah data mentah menjadi kumpulan data yang sudah siap digunakan [15]. Berikut adalah beberapa tahapan pre-processing yang dilakukan oleh penelitian ini:

- Case folding* : tahapan yang mengubah setiap huruf dalam dokumen menjadi huruf kecil atau *lower case* tetapi hanya huruf "a" hingga "z" yang di ubah menjadi *lower case*.
- Cleaning* : tahapan mengurangi kebisingan selama proses klasifikasi dengan membersihkan review kata-kata yang tidak diperlukan. Semua karakter non-abjad dihapus.
- Tokenize* : proses pemecahan kata menjadi kata-kata tunggal dilakukan dengan memindai data dan setiap kata teridentifikasi dan di pisahkan dengan kata yang lain oleh pemisah spasi.
- Stopword Removal* : daftar kata akan dibersihkan guna meningkatkan ketepatan klasifikasi.
- Stemming* : proses untuk mengidentifikasi stem atau kata dasar dari setiap kata dalam kalimat dengan membedakan kata-kata tersebut dari imbuhan seperti prefiks dan sufiks.
- Pembobotan kata : dalam proses di lakukan pembobotan kata dengan TF-IDF untuk menilai relevansi kata atau frasa dalam dokumen.
- Labeling* : Pada tahap ini mempunyai dua kategori label yaitu positif dan negatif. Pada kategori positif yaitu dengan ulasan yang memiliki *score* 3 sampai *score* 5 sedangkan untuk ulasan negatif memiliki *score* 1 dan *score* 2 dan di validasi oleh pakar bahasa indonesia.

2.1.4 Word Cloud

Word Cloud merupakan gambar visual yang menunjukkan berapa kali kata-kata muncul dalam kumpulan teks. Frekuensi kata ditentukan oleh ukuran huruf, semakin besar ukuran huruf, semakin banyak kata yang muncul. Sebaliknya, semakin kecil ukuran huruf, semakin sedikit kata yang muncul [16].

2.1.5 Implementasi algoritma

Pada tahap berikutnya, data yang sudah siap diolah digunakan untuk memodelkan klasifikasi teks. Sebelum di implementasikan algoritma, data harus melalui tahapan pembobotan TF-IDF, metode yang diterapkan dalam analisis teks untuk mengukur tingkat kepentingan setiap kata atau frasa dalam suatu dokumen [17]. Kemudian data dibagi menjadi 80% data uji dan 20% data latih. Selanjutnya, pemodelan dilakukan menggunakan algoritma Logistic regression dan K-nearest neighbor.

a. Algoritma Logistic Regression

Teknik Logistic Regression merupakan sebuah metode statistik yang bekerja berdasarkan hubungan antara variabel satu atau lebih [18]. Probabilitas adalah suatu peristiwa akan terjadi dari banyaknya peristiwa dalam suatu percobaan. Probabilitas selalu berkisar antara 0 dan 1 [19].

$$O(Y) = \frac{p^{(y)}}{1-p^{(y)}} \quad (1)$$

Odds Ratio (O(Y)) adalah perbandingan antara probabilitas kejadian negatif dan probabilitas kejadian positif dari suatu variabel Y. Dalam rumus ini, P(Y) adalah probabilitas kejadian positif, sementara 1-P(Y) adalah probabilitas kejadian negatif. Odds Ratio membantu untuk memahami seberapa besar kemungkinan suatu kejadian negatif terjadi dibandingkan dengan kejadian positif.

b. Algoritma K-Nearest Neighbor

K-Nearest Neighbor (KNN) termasuk kelompok instance-based learning. Algoritma ini juga merupakan salah satu teknik lazy learning. KNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing [20]. Jarak Euclidean dapat dicari dengan menggunakan persamaan pada langkah-langkah algoritma KNN berikut :

$$d_i \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Menghitung jarak Euclidean antara dua vektor dalam ruang n-dimensi. Di sini, x_i adalah komponen dari data training dan y_i adalah komponen dari data testing. Rumus ini menggunakan akar kuadrat dari jumlah kuadrat selisih antara setiap pasangan komponen dari kedua vektor. Parameter n adalah jumlah data atau dimensi dari vektor yang digunakan dalam perhitungan ini.

2.1.6 Evaluasi

Data yang sudah di lakukan penerapan algoritma Kemudian di evaluasi, tujuannya adalah untuk mengevaluasi model model algoritma klasifikasi yang telah di terapkan. Evaluasi ini menggunakan confusion matrix dengan mengukur akurasi (Accuracy) prediksi dengan mempertimbangkan data secara keseluruhan, baik positif maupun negatif, perbandingan antara prediksi yang benar dengan prediksi yang benar positif secara keseluruhan (Precision), prediksi

tidak bisa login hanya loading saja [‘tidak’, ‘bisa’, ‘login’, ‘hanya’, ‘loading’, ‘saja’]

d. *Stemming*

Proses ini dilakukan penghapusan semua kata imbuhan yang ada pada kata menjadi kata dasar yang tersedia di kamus besar bahasa Indonesia. Sebagai contoh imbuhan yaitu : meng-,me-,kan-,kan-,di-, dan sejenisnya. Dapat di lihat pada tabel 4.

Table 4. Hasil Stemming

Sebelum	Sesudah
[‘berguna’, ‘bangat’]	Guna bangat
[‘gimana’, ‘cara’, ‘mengatasi’, ‘kesalahan’, ‘koneksi’]	gimana cara atasi salah koneksi

e. *Stopword Removal*

Pada tahap ini melakukan penghapusan dan penyaringan kata pada data teks. Kata yang di hapus berupa kata yang di rasa tidak cukup penting dan tidak berpengaruh pada analisis sentimen ini. Dapat dilihat pada tabel 5.

Table 5. Hasil *Stopword Removal*

Sebelum	Sesudah
Aplikasi sampah dah x download ttp ga bsa dibuka	Aplikasi sudah download tetap ga bisa di buka

f. Pembobotan kata

Tahap pembobotan kata ini digunakan untuk menghitung nilai kata-kata dalam teks yang bertujuan untuk mendapatkan nilai masing masing kata yang akan di gunakan sebagai fitur. Tabel 6 menunjukkan hasil pembotoan kata.

Table 6. Hasil pembobotan kata

Term	TF	IDF	TF-IDF
Capil	2	5.067676315	0.7710726739
Wifi	2	3.822460552	0.5816067751
Barcodenya	2	6.625820933	0.9674052731
Koneksi	1	2.84732932	0.3141852107
Masuk	5	3.014903021	0.5795384362
Gagal	1	4.48575477	0.8299608119

g. *Labelling*

Setelah dataset sudah di dapatkan kemudian data set tersebut di proses untuk labeling sentimen menjadi dua label yaitu positif dan negatif, dimana nilai negatif memiliki score ulasan 1-2, nilai positif 3-5 dan juga di validasi oleh pakar bahasa indonesia. Pada gambar 5 penjelasan mengenai dataset hasil validasi oleh pakar bahasa indonesia.

Table 7. Hasil *Labelling* Pakar

Text	Sentimen	Correct
tolong saya paspor saya tablokir	Positif	Positif
harus pakai wifi	Negatif	Positif
sangat bermanfaat	Positif	Positif
login nya sulit	Positif	Negatif
ribed banget udah pakai apk masih tetap harus datang ke dukcapil bikin nambah krjaan	Positif	Negatif
sayang sekali ini cuma bisa di gunakan di jabodetabek	Negatif	Positif

3.3 WordCloud

Hasil pengolahan data di representasikan menggunakan *worldcloud*. Setiap label sentiment negatif dan positif pada data di visualisasikan menggunakan *wordcloud*. Gambar dan gambar 3 menunjukkan hasil dari *worcloud* sentimen positif dan negatif.



Gambar 3. Hasil *wordcloud* sentimen positif

Berdasarkan visualisasi wordcloud diatas di ketahui kata-kata yang memiliki frekuensi paling sering muncul terbanyak yaitu daftar,urus,masuk,data. Gambar 3 menunjukkan wordcloud ulasan sentimen positif yang sering muncul dengan tampilan kata yang terbesar.

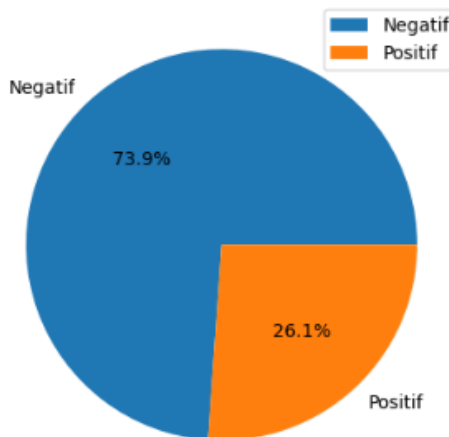


Gambar 4. Hasil wordcloud sentimen negatif

Berdasarkan visualisasi wordcloud diatas di ketahui kata-kata yang memiliki frekuensi paling sering muncul terbanyak yaitu koneksi,wifi,data. Gambar 4 menunjukkan wordcloud ulasan sentimen negatif yang sering muncul dengan tampilan kata yang terbesar.

3.4 Klasifikasi Algoritma

Setelah tahapan pre-processing selesai tahap berikutnya dilakukan klasifikasi. Tahap ini di lakukan mengguna code editor google collab menggunakan bahasa pemrograman python dan menggunakan algoritma logistic-regression dan k-nearest neighbor. Pada tahap ini data yang di gunakan berjumlah 1108 data yang sudah di pre-processing dan memiliki 2 label sentimen yaitu positif dan negatif yang telah di validasi oleh pakar bahasa indonesia. Gambar 5 menunjukkan hasil perbandingan label sentimen positif dan negatif.



Gambar 5. Perbandingan sentimen negatif dan positif

Gambar 5 menjelaskan bahwa kelas positif mendapatkan jumlah sebanyak 73.9% data sedangkan kelas negatif sebanyak 26.1% data. Kemudian, data dipisahkan menjadi data training dan data testing dengan ratio 0.2 atau data training 80% sebanyak data dan data testing 20% sebanyak 222 data. Data yang di gunakan untuk mencari nilai akurasi menggunakan algoritma adalah data testing. Berikut tabel 8 menunjukkan hasil performance kedua algoritma.

Table 8. Hasil Performance algoritma

NO	Algoritma	Akurasi(%)	Precision(%)	Recall(%)	F1-Score(%)
1	Logistic regression	78.83%	71.63%	55.84%	55.34%
2	K-nearest Neighbor	82.43%	77.88%	65.97%	68.72%

Berdasarkan pada tabel 8 diatas algoritma logistic regression mendapatkan nilai accuracy 78.83%, recall 55.84%, precision 71.63% dan f1-score 55.34% dan algoritma k-nearest neighbor mendapatkan nilai accuracy 82.43%, recall 65.97%, precision 77.88% dan f1-score 68.72%. dan dapat disimpulkan bahwa nilai performa akurasi dari model algoritma k-nearest neighbor lebih baik dengan nilai akurasi 82.43% dengan algoritma logistic regression dengan selisih 3.60% menjadi algoritma k-nearest neighbor lebih baik dari algoritma logistic regression.

3.5 Evaluasi

Tahap evaluasi ini menggunakan confusion matrix untuk mengetahui bagaimana klasifikasi masing-masing algoritma bekerja pada jumlah data label yang diprediksi. Dan juga melakukan penghitungan ulang pada hasil akurasi yang di peroleh setiap algoritma.



Gambar 6. Confusion matrix logistic regression

Berdasarkan gambar 6 menunjukkan bahwa jumlah 168 True Positive, 43 False Positive, 4 False Negative dan 7 True Negative. Setelah itu di hitung dengan $TP+FP+FN+TN = 222$ Kemudian $(TP+TN)/222 = 0.788$. hasil dari perhitungan ulang dengan confusion matrix Logistic regression adalah 78.8%.



Gambar 7. Confusion matrix k-nearest neighbor

Berdasarkan gambar 7 menunjukkan bahwa jumlah 165 True Positive, 32 False Positive, 7 False Negative dan 18 True Negatif . Kemudian di hitung dengan $TP+FP+FN+TN = 222$ Kemudian $(TP+TN)/222 = 0.824$. hasil dari perhitungan ulang dengan confusion matrix K-Nearest Neighbor adalah 82.4%.

4. KESIMPULAN

Kesimpulan dari proses analisis aplikasi identitas kependudukan digital adalah bahwa aplikasi ini masih berkualitas rendah dalam penggunaannya oleh masyarakat dan masyarakat masih merasa kesulitan dalam menggunakan aplikasi identitas kependudukan digital karena memperoleh ulasan negatif sebesar 73.9% atau 819 ulasan negatif dan ulasan positif sebesar 26.1% atau 289 ulasan positif terhadap ulasan aplikasi identitas kependudukan digital. Oleh karena itu, aplikasi identitas kependudukan digital harus diperbaiki agar lebih efisien dalam menyediakan layanan data kependudukan masyarakat dan membuat penggunaanya puas. Adapun hal lainnya yaitu hasil pengujian untuk membandingkan algoritma *Logistic Regression* dan *K-Nearest Neighbor* serta mengetahui performa terbaik kedua algoritma. Hasil penelitian menunjukkan bahwa metode *K-nearest neighbor* Adalah pendekatan paling optimal untuk menganalisis sentimen terkait ulasan aplikasi identitas kependudukan digital dengan tingkat akurasi mencapai 82.43%, presisi sebesar 77.88%, dan recall sebesar 65.97%. Data dapat digunakan sebagai rekomendasi pengembangan untuk penelitian berikutnya yang terbaru dan memperbanyak data lalu mempertimbangkan teknik pengolahan data yang lebih baik untuk meningkatkan performa dan kinerja algoritma.

REFERENCES

- [1] E. Putri Primawanti and H. Ali, "Pengaruh Teknologi Informasi, Sistem Informasi Berbasis Web Dan Knowledge Management Terhadap Kinerja Karyawan (Literature Review Executive Support Sistem (Ess) for Business)," *J. Ekon. Manaj. Sist. Inf.*, vol. 3, no. 3, pp. 267–285, 2022, doi: 10.31933/jemsi.v3i3.818.
- [2] F. Parsakh Nursyamsyi and F. Noor Hasan, "KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Sentimen Terhadap Aplikasi Identitas Kependudukan Digital Menggunakan Algoritma Naïve Bayes dan SVM," *Media Online*, vol. 4, no. 3, pp. 1788–1798, 2023, doi: 10.30865/klik.v4i3.1517.
- [3] N. Wahyuningsih and Hendry, "Perbandingan Metode Klasifikasi Dalam Analisis Sentimen Masyarakat Terhadap Identitas Kependudukan Digital (Ikd)," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 8, no. 4, pp. 1218–1227, 2023.
- [4] R. Hidayat, R. Nur Rahman, M. Reifin Perdana, and Arbansyah, "Analisis Sentimen Aplikasi Identitas Kependudukan Digital



- (IKD)Menggunakan Metode Naïve Bayes,” *J. Sist. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 129–140, 2024, [Online]. Available: <https://doi.org/10.59581/jusiik-widyakarya.v2i1.2320>
- [5] D. Prاتمanto, F. F. D. Imaniawan, and V. Maarif, “Analisis Sentimen Pada Ulasan Pengguna Aplikasi Identitas Kependudukan Digital Dengan Metode Naive Bayes Dan K-Nearest,” *Comput. J. Comput. Sci. Inf. Syst.*, vol. 7, no. 2, pp. 155–166, 2023, doi: 10.24912/computatio.v7i2.26322.
- [6] D. D. PURNAMASARI, “pengguna Identitas Kependudukan Digital Ditarget Capai 50 Juta Jiwa di 2023,” 2023. <https://www.kompas.id/baca/polhuk/2023/05/03/pengguna-identitas-kependudukan-digital-ditarget-capai-50-juta-jiwa-di-2023> (accessed Jan. 10, 2024).
- [7] Sri Handriana Dewi Hastuti, “Pentingnya Pemanfaatan Data Kependudukan Di Era Digital,” *Tek. Teknol. Inf. dan Multimed.*, vol. 1, no. 1, pp. 18–21, 2020, doi: 10.46764/teknimedia.v1i1.9.
- [8] M. R. Ramli, “Analisis Sentimen Terhadap Opini Mahasiswa Terakit Pembelajaran Daring Dengan Menggunakan Algoritma Naive Bayes Classifier dan SVM dengan Adaboost Pada media Sosial Twitter,” Universitas Siliwangi Tasikmalaya, 2022. [Online]. Available: <http://repositori.unsil.ac.id/8460/>
- [9] V. A. Permadi, “Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura,” *J. Buana Inform.*, vol. 11, no. 2, pp. 141–151, 2020, doi: 10.24002/jbi.v11i2.3769.
- [10] R. A. Lestari *et al.*, “penerapan Algoritma Support Vector Machine Pada Analisis Application Of Suppot Vector Machine Algorithm In Sentiment,” *J. Tek. inform. dan ilmu komput.*, vol. 10, no. 5, pp. 1063–1070, 2023, doi: 10.25126/jtiik.2023107264.
- [11] R. AL Anshari *et al.*, “Menggunakan Metode Support Vector Machine 1) Info Artikel Abstrak,” *J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 118–128, 2023, doi: 10.55123.
- [12] M. Raja Nurhusen, J. Indra, and K. Ahmad Baihaqi, “JURNAL MEDIA INFORMATIKA BUDIDARMA Analisis Sentimen Pengguna Twitter Terhadap Kenaikan Harga Bahan Bakar Minyak (BBM) Menggunakan Metode Logistic Regression,” *J. Media Inform. Budidarma*, vol. 7, no. 1, pp. 276–282, 2023, doi: 10.30865/mib.v7i1.5491.
- [13] S. Nurhaliza, Y. Yusra, and M. Fikry, “Klasifikasi Sentimen Masyarakat di Twitter Terhadap Kenaikan Harga BBM dengan Metode Support Vector Machine,” *J. Sist. Komput. dan Inform.*, vol. 4, no. 4, p. 586, 2023, doi: 10.30865/json.v4i4.6322.
- [14] A. Maulana, I. K. Afifah, A. Mubarrak, and K. R. Fauzan, “Comparison of Logistic Regression , Multinomialnb , Svm , and K-Nn Methods on Sentiment Analysis of Gojek App Reviews on the Google Play Store Perbandingan Metode Logistic Regression , Multinomialnb , Svm , Dan K-Nn Pada Analisis Sentimen Ulasan Aplikasi,” *J. Tek. inform.*, vol. 4, no. 6, pp. 1487–1494, 2023, doi: 10.52436/1.jutif.2023.4.6.863.
- [15] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [16] R. Parluka, S. I. Pradika, A. M. Hakim, and K. R. N M, “Analisis Sentimen Twitter Terhadap Bitcoin Dan Cryptocurrency Berbasis Python Textblob,” *J. Ilm. Teknol. Inf. dan Robot.*, vol. 2, no. 2, pp. 33–37, 2020, doi: 10.33005/jifti.v2i2.22.
- [17] V. Dwi Antonio, S. Efendi, and H. Mawengkang, “Sentiment Analysis for Covid-19 in Indonesia on Twitter with TF-IDF Featured Extraction and Stochastic Gradient Descent,” *Int. J. Nonlinear Anal. Appl.*, vol. 13, no. 1, pp. 2008–6822, 2022, [Online]. Available: <http://dx.doi.org/10.22075/ijnaa.2021.5735>
- [18] B. Ramadhani and ryan randy Suryono, “Komparasi Algoritma Naive Bayes dan Logistic regression Untuk Analisis Sentimen Metaverse,” *J. Mnemon.*, vol. 7, no. 1, pp. 31–39, 2024, doi: 10.36040/mnemonic.v7i1.8977.
- [19] M. I. Gunawan, D. Sugiarto, and I. Mardianto, “Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode Grid Search pada Algoritma Logistic Regression,” *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 280, 2020, doi: 10.26418/jp.v6i3.40718.
- [20] H. Leidiyana, “Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor,” *J. Penelit. Ilmu Komputer, Syst. Embed. Log.*, vol. 1, no. 1, pp. 65–76, 2020.