



Analisis Sentimen Pengguna Media Sosial Terhadap Identitas Kependudukan Digital Menggunakan Metode Support Vector Machine (SVM)

Vina Alipah Sulistiani, Muhammad Hamka*

Fakultas Teknik dan Sains, Teknik Informatika, Universitas Muhammadiyah Purwokerto, Banyumas
Jl. KH. Ahmad Dahlan, Dusun III, Dukuhwaluh, Kec. Kembaran, Kabupaten Banyumas, Jawa Tengah, Indonesia

Email: ¹vinaalipah.s@gmail.com, ^{2,*}muhammadhamka@ump.ac.id

Email Penulis Korespondensi: muhammadhamka@ump.ac.id

Submitted: 17/07/2024; Accepted: 31/07/2024; Published: 31/07/2024

Abstrak—Pemerintah Republik Indonesia telah mengimplementasikan kebijakan penggunaan Identitas Kependudukan Digital (IKD) sebagai salah satu layanan prioritas Sistem Pemerintahan Berbasis Elektronik (SPBE) untuk percepatan transformasi digital dalam rangka integrasi layanan digital nasional. Implementasi IKD menimbulkan perspektif pro dan kontra di masyarakat. Perspektif kontra yang merupakan permasalahan dalam penggunaan IKD yaitu masyarakat masih khawatir akan keamanan dan privasi data diri yang tercantum pada sistem digital. Masyarakat dapat dengan mudah menyuarakan opini, aspirasi, dan partisipasi melalui media sosial. Namun, mengingat banyaknya data media sosial yang ada, proses memperoleh informasi yang tepat memerlukan banyak waktu dan keterampilan khusus. Sehingga dilakukan analisis sentimen terhadap IKD pada media sosial menggunakan metode Support Vector Machine (SVM) dan ditingkatkan dengan teknik Synthetic Minority Oversampling Technique (SMOTE) yang bertujuan untuk mengetahui pandangan atau pendapat terhadap IKD yang bersifat positif, negatif, dan netral berdasarkan sudut pandang orang lain. Data opini yang digunakan dalam penelitian ini sebanyak 6697 dengan kata kunci “identitas kependudukan digital”, “ktp elektronik”, dan “ktp digital” yang diambil dari tahun 2021 sampai tahun 2024. Kemudian, data diklasifikasikan berdasarkan nilai polaritas menggunakan kamus lexicon yaitu senticnet7 dengan hasil sentimen positive sebesar 57,1% dan sentimen negative sebesar 42,9%. Hasil dari proses klasifikasi menggunakan metode SVM dan ditingkatkan dengan teknik SMOTE mempunyai hasil accuracy sebesar 89%, presisi negatif sebesar 88%, presisi positif sebesar 90%, recall negatif sebesar 85%, recall positif sebesar 92%, f1-score negative sebesar 86%, dan f1-score positive sebesar 91%.

Kata Kunci: Identitas Kependudukan Digital; Analisis Sentimen; SVM; SMOTE

Abstract—The Government of the Republic of Indonesia has implemented a policy of using Digital Population Identity (IKD) as one of the priority services of the Electronic Based Government System (SPBE) to accelerate digital transformation in the context of integrase national digital service. The implementation of IKD raises pros and cons perspectives in society. The counter perspective that is a problem in using IKD is that people are still worried about the security and privacy of personal data listed on digital systems. People can easily voice opinions, aspirations and participation through social media. However, given the large amount of social media data available, the process of obtaining the right information requires a lot of time and special skills. So a sentiment analysis of IKD on social media was carried out using the method Support Vector Machine (SVM) and improved with techniques Synthetic Minority Oversampling Technique (SMOTE) which aims to find out views or opinions regarding IKD which are positive, negative and neutral based on other people's points of view. The opinion data used in this research was 6697 with the keywords "digital population identity", "electronic ID card", and "digital ID card" taken from 2021 to 2024. Then, the data was classified based on polarity values using a dictionary lexicon that is senticnet7 with sentiment results positive by 57.1% and sentiment negative amounting to 42.9%. The results of the classification process using the SVM method and enhanced with the SMOTE technique have results accuracy of 89%, negative precision of 88%, positive precision of 90%, recall negative by 85%, recall positive by 92%, f1-score negative by 86%, and f1-score positive by 91%.

Keywords: Digital Population Identity; Sentiment Analysis; SVM; SMOTE

1. PENDAHULUAN

Pemerintah Republik Indonesia telah mengimplementasikan kebijakan penggunaan Identitas Kependudukan Digital (IKD) sebagai salah satu layanan prioritas Sistem Pemerintahan Berbasis Elektronik (SPBE) untuk percepatan transformasi digital dalam rangka integrasi layanan digital nasional [1]. Manfaat adanya Identitas Kependudukan Digital adalah mempermudah dan mempercepat transaksi pelayanan publik dalam bentuk digital serta mengamankan kepemilikan identitas kependudukan digital melalui sistem autentikasi untuk mencegah pemalsuan data [2].

Implementasi IKD menimbulkan perspektif pro dan kontra di masyarakat. Masyarakat yang setuju dengan penerapan IKD berpendapat bahwa masyarakat tidak perlu lagi membawa KTP fisik untuk mengecek data diri bila dibutuhkan, cukup membawa ponsel yang sudah terinstall aplikasi IKD [3]. Sedangkan pandangan yang tidak sepakat dengan kebijakan penggunaan IKD terkait aspek keamanan dan privasi data pribadi yang tercantum dalam sistem digital [4]. Untuk menyelesaikan permasalahan tersebut, pemerintah harus melakukan upaya dengan cara sosialisasi kepada masyarakat secara berkala terkait penggunaan IKD dan meningkatkan keamanan dan privasi data masyarakat [5]. Opini masyarakat memiliki peran penting dalam mempengaruhi kebijakan pemerintah untuk memperhatikan kepentingan dan membutuhkan dukungan negara guna membantu mempercepat akses publik terhadap layanan digital yang disediakan oleh pemerintah [6]. Masyarakat dapat dengan mudah untuk menyuarakan pendapat, aspirasi, dan partisipasi dalam perdebatan publik terhadap isu-isu terkini [7]. Menganalisis

data sumber pengambilan keputusan dapat dilakukan melalui media sosial, namun mengingat banyaknya data media sosial yang ada, proses memperoleh informasi yang tepat memerlukan banyak waktu dan keterampilan khusus [8]. Maka dari itu, dibutuhkan analisis sentimen pendapat media sosial yang bertujuan untuk mengetahui pandangan atau pendapat terhadap suatu isu berdasarkan sudut pandang orang lain, dengan mempertimbangkan apakah sudut pandang isu itu bersifat positive, negative, atau netral [9].

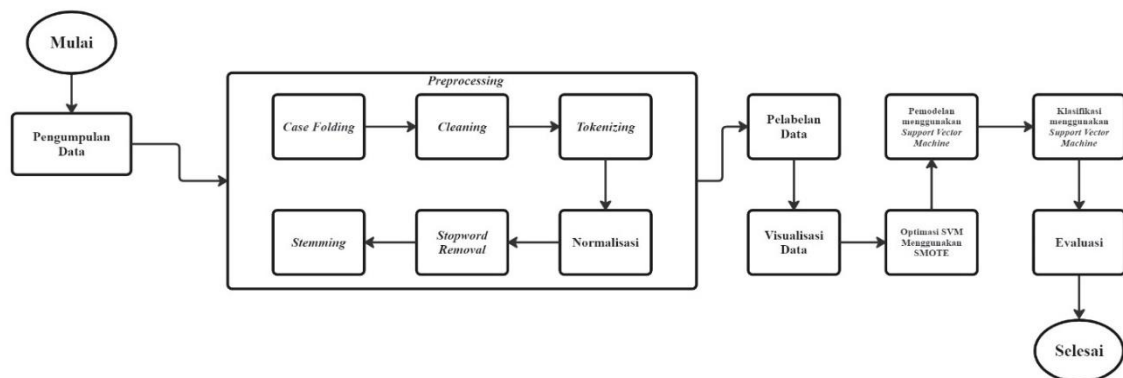
Analisis sentimen dapat dilakukan melalui beberapa metode. Studi yang dilakukan oleh [10] metode Support Vector Machine (SVM) IKD digunakan untuk analisis sentimen dan menghasilkan skor sebesar 78,27%, negatif sebesar 12,97% netral dan positif sebesar 8,76% dengan menggunakan perbandingan 80: 20 antara data latih dan uji, maka akurasi data uji yang dihasilkan oleh SVM adalah 77%. Kajian analisis sentimen lain dilakukan oleh [11] menggunakan metode Naïve Bayes terhadap IKD menghasilkan akurasi, presisi, recall, dan F1 sebesar 90%. Studi analisis sentimen yang dilakukan oleh [12] menggunakan metode SVM terhadap aplikasi jamsostek mobile dengan pendekatan kernel linier menghasilkan nilai akurasi terbaik dengan accuracy sebesar 96%, presisi sebesar 92%, recall sebesar 96%, dan f1-score sebesar 94%, sedangkan untuk hasil komentar terbanyak adalah komentar berkategori positive dengan jumlah sebesar 17.571. Kajian analisis sentimen lain dilakukan oleh [13] menggunakan klasifikasi SVM Gojek di media sosial Twitter, diperoleh akurasi keseluruhan sebesar 79,19% dan akurasi Kappa sebesar 21%. Semakin tinggi nilai akurasi keseluruhan dan akurasi kappa yang diperoleh, semakin baik kinerja model klasifikasi. Penelitian lainnya memanfaatkan Synthetic Minority Over-sampling Technique (SMOTE) dalam analisis sentimen pada metaverse menunjukkan bahwa metode Logistic Regression mempunyai nilai akurasi yang cukup baik yaitu sebesar 95%. Kemudian terjadi peningkatan pada confusion matrix dengan nilai presisi sebesar 94%, recall sebesar 93% dan F1-Score sebesar 95%. Selanjutnya pada metode Naïve Bayes memiliki nilai yang lebih kecil dari metode sebelumnya yaitu dengan nilai accuracy sebesar 91%, confusion matrix dengan nilai precision negatif sebesar 87%, recall sebesar 97% dan F1-Score sebesar 92%, sehingga hasil yang memiliki nilai performa lebih baik yaitu nilai accuracy dan confusion matrix [14]. Keterbatasan dalam penelitian analisis sentimen menggunakan metode SVM yaitu ketergantungan pada ekstraksi fitur yang baik untuk menghasilkan kinerja yang optimal, jika fitur yang diekstraksi tidak mampu menangkap nuansa semantik dari teks, akurasi model dapat menurun [15]. Maka, perbaikan analisis sentimen pada metode SVM yaitu penggunaan ekstraksi yang lebih canggih yang dapat meningkatkan kualitas fitur yang digunakan oleh SVM [16]. Alasan menggunakan SVM dalam perbaikan proses analisis sentimen yaitu kemampuan generalisasi yang baik terutama ketika hyperparameter dioptimalkan dengan baik [17].

Artikel ini akan membahas pemodelan untuk klasifikasi sentimen terkait penerapan IKD. Algoritma yang digunakan adalah SVM. Selain itu, digunakan juga metode SMOTE untuk mengatasi ketidakseimbangan kelas pada sentimen positif dan negatif tidak seimbang. SMOTE bekerja dengan cara pembuatan data secara acak pada kelas minoritas [18]. Pembuatan data sintesis melalui proses interpolasi [19], yaitu memperkirakan pembuatan data yang tidak diketahui pada suatu kelas, sehingga akan terciptanya fungsi data yang kontinu [20].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian yang digunakan dalam analisis sentimen terdiri dari pengumpulan data, preprocessing yang meliputi case folding, cleaning, tokenizing, normalisasi, stopword removal, stemming. Fase berikutnya adalah pelabelan data, visualisasi data, optimasi SVM melalui SMOTE. Langkah selanjutnya pembuatan model klasifikasi menggunakan SVM, kemudian dilanjutkan evaluasi model. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

2.2 Pengumpumpulan Data

Pengumpulan data menggunakan opini media sosial X dengan teknik crawling dan bahasa pemrograman python serta google colab [21] [22]. Tweet yang digunakan merupakan komentar media sosial X dari tahun 2021 sampai



2024. Adapun kata kunci yang digunakan adalah identitas kependudukan digital [23], ktp elektronik [24], dan ktp digital [3].

2.3 Preprocessing

Preprocessing adalah tahapan untuk proses pengolahan data dengan cara melakukan pembersihan atau cleaning dan mengatur suatu data mentah menjadi data yang bersih dan teratur. Pembersihan data adalah tahapan penting dalam masalah text mining yang akan dianalisis, karena biasanya penuh dengan noise yang mempengaruhi kinerja analisis [25]. Tahap Preprocessing pada penelitian analisis sentimen ini terdiri dari case folding, cleaning, tokenizing, normalisasi, stopword removal, dan stemming.

1. Case folding merupakan tahap merubah huruf kapital menjadi huruf kecil atau disebut dengan lowercase yang bertujuan untuk memudahkan pencarian serta menghindari kalimat sensitif dari proses klasifikasi [26].
2. Cleaning merupakan tahap pembersihan data dari tanda baca, simbol, atau elemen yang tidak relevan dalam penulisan [27].
3. Tokenizing merupakan tahap membagi teks menjadi bagian-bagian kata yang terstruktur sehingga dapat menghitung jumlah frekuensi, seperti “saya pergi ke pasar” menjadi “saya”, “pergi”. “ke”, “pasar” [28].
4. Normalisasi merupakan tahap memperbaiki kata-kata yang singkat dan tidak baku menjadi bentuk kata baku sesuai dengan kamus yang digunakan dalam normalisasi [28].
5. Stopword removal merupakan tahap menghapus kata yang tidak dibutuhkan dalam analisis dan kemunculannya hanya akan memperburuk performa [29].
6. Stemming merupakan tahap mengubah kata yang mengandung imbuhan pada data tweet menjadi kata dasar. Proses stemming dilakukan dengan menggunakan bahasa pemrograman python sastrawi. [29].

2.4 Pelabelan Data

Pelabelan data bertujuan untuk memberikan label sentimen pada opini. Label dikelompokkan ke dalam sentimen positive, negative, dan netral. Proses ini tidak dapat dilakukan secara manual dengan melihat opini satu persatu [30]. Oleh karena itu, perlu adanya suatu proses pelabelan sentimen dengan menggunakan metode berbasis lexicon-based method. Kamus yang digunakan adalah senticnet 7 [31]. Pelabelan dilakukan dengan membuat fungsi yang menghitung nilai polaritas dari suatu kumpulan opini berdasarkan Persamaan (1). Penentuan nilai polaritas berdasarkan skor opini menggunakan persamaan (2) [8].

$$\text{Score} = \sum_{i=1}^n \text{Score} + W_{\text{positive}} + W_{\text{negative}} \quad (1)$$

Persamaan (1) diperoleh dari hasil penjumlahan rata-rata setiap kata positive dan negative yang masuk ke dalam kamus lexicon [32]

$$\text{Score} \begin{cases} \text{Positive, jika score} > 0 \\ \text{Netral, jika score} = 0 \\ \text{Negative, jika score} < 0 \end{cases} \quad (2)$$

Persamaan (2) adalah nilai polaritas dengan kriteria bahwa sebuah opini dianggap memiliki sentimen positive jika nilai polaritas lebih dari nol (>0), jika nilai polaritas sama dengan nol ($= 0$) maka dianggap memiliki sentimen netral, dan jika memiliki nilai polaritas kurang dari nol (< 0) maka dianggap memiliki sentimen negative [32].

2.5 Visualisasi Data

Penelitian ini menggunakan visualisasi data berupa wordcloud dan pie chart [33]. Wordcloud untuk menunjukkan visualisasi frekuensi kata dalam kumpulan teks yang sering muncul. Besar kecilnya huruf menentukan frekuensi kemunculan sebuah kata, semakin besar ukuran huruf maka semakin besar pula kemunculan kata tersebut [34]. Sedangkan pie chart digunakan untuk memvisualisasikan distribusi data [34]. Visualisasi data membantu pengguna dalam menganalisis informasi dan mendapatkan pengetahuan tentang opini IKD [35].

2.6 Optimasi SVM melalui SMOTE

Optimasi SVM menggunakan SMOTE digunakan untuk mengubah data yang tidak seimbang dengan membuat data sintesis baru pada kelas minoritas yang bertujuan untuk meningkatkan kinerja dalam metode klasifikasi [36]. Metode SMOTE bekerja dengan cara memilih sampel secara acak dari setiap kelas minoritas dan membuat sampel sintesis yang baru dengan cara melakukan interpolasi antara sampel minoritas yang dipilih dengan menggabungkan fitur dari kedua sampel. Hasil sampel sintesis akan ditambahkan ke data pelatihan dan akan meningkatkan representasi data kelas minoritas dan membantu mengurangi bias model terhadap kelas mayoritas [37].

2.7 Pemodelan Support Vector Machine (SVM)

Pemodelan SVM dalam penelitian ini melibatkan hyperparameter. Parameter yang digunakan yaitu kernel linear. Data yang telah diolah akan diproses lanjut menggunakan library Natural Language Toolkit (NLTK). Proses



pembuatan model SVM yaitu dengan memasukkan data uji yang telah diambil, sehingga data yang telah terkumpul menunjukkan hasil nilai pada setiap prediksi yang dibuat dengan model tersebut [30] [38].

2.8 Klasifikasi Menggunakan Support Vector Machine (SVM)

Klasifikasi pada tahap ini digunakan untuk menentukan sentimen dari opini tweet. SVM mempunyai prinsip dasar klasifikasi linier yaitu klasifikasi yang dapat dipisahkan secara linier, namun SVM dikembangkan untuk mampu menyelesaikan permasalahan non-linier dengan menerapkan konsep kernel pada ruang kerja berdimensi tinggi. Pada ruang berdimensi tinggi, akan mencari hyperplane yang memaksimalkan jarak (margin) antar kelas data. [39]. Rumus Perhitungan algoritma SVM dengan kernel linear sesuai dengan persamaan (3) [39]:

$$f(x) = w \cdot x + b \tag{3}$$

Keterangan : f = Nilai target, (x) = vector input, w = vector bobot, b = bias

2.9 Evaluasi Model

Evaluasi digunakan untuk mengetahui akurasi dari model metode yang telah dibuat. Pada penelitian ini alat yang digunakan untuk menganalisis model adalah confusion matrix untuk menganalisis seberapa baik model klasifikasi mengenali tuple dari data yang berbeda. Pengujian dilakukan dengan menggunakan confusion matrix melalui library SVM, dari model pelatihan dan pengujian yang telah diklasifikasikan sebelumnya.

Terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi dalam pengukuran kinerja menggunakan confusion matrix, antara lain True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) [40].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian analisis sentimen ini diambil dari opini masyarakat pada media sosial X dengan teknik crawling menggunakan bahasa pemrograman python pada google colab. Diambil dari tahun 2021 sampai tahun 2024 sebanyak 6697 opini masyarakat. Ditunjukkan dalam Tabel 1.

Tabel 1. Hasil Pengumpulan Data

| No | Hasil Pengumpulan Data |
|------|---|
| 1 | @AdhiWeecak Selamat pagi Kak. Penumpang KA usia 17 tahun atau lebih saat boarding wajib menunjukkan identitas aktif asli dan berfoto. Silakan dapat menunjukkan KTP digital dan wajib dari aplikasi Dukcapil (Dinas Kependudukan dan Pencatatan Sipil) yang ter |
| 2 | @KAI121 apakah bisa check in KAJJ dengan menggunakan KTP yg tersimpan dalam aplikasi Identitas Kependudukan Digital yang mana aplikasi tersebut adalah aplikasi resmi dari Kementerian Dalam Negeri? |
| ... | ... |
| 6696 | @kompascom Nah bagus ini. Kedepannya tidak ada lagi namanya fotokopi ktp. Beneran primitif gais kalo kita apa masih harus fotokopi. Di negara maju semua serba digital m cardless. satu data bisa diakses untuk kebutuhan apapun. #ktpdigital |
| 6697 | Kota Jambi akan Terapkan KTP Digital Kadis Dukcapil: Tak Perlu Cetak https://t.co/JZaIxLgNYf via @tribunjambiku #ktp #ktpdigital #kotajambi #dukcapil |

3.2 Preprocessing

Kumpulan opini-opini yang sudah di kumpulkan dengan crawling akan diolah data dengan cara melakukan cleaning dan mengatur suatu data mentah menjadi data yang bersih dan terstruktur. Tahap preprocessing pada penelitian analisis sentimen ini meliputi case folding, cleaning, tokenizing, normalisasi, stopword removal, dan stemming.

a. Case Folding

Case folding dilakukan untuk mengubah semua huruf menjadi huruf kecil. Pada proses ini huruf 'A' - 'Z' yang terdapat pada data diubah mrnjjadi huruf 'a' - 'z'. Hasil case folding ditunjukkan pada Tabel 2.

b. Cleaning dilakukan untuk mengurangi noise dari data case folding. Hasil ditunjukkan pada Tabel 2.

c. Tokenizing dilakukan untuk memisahkan kalimat menjadi kata per kata atau memisahkan urutan string menjadi potongan kata. Hasil ditunjukkan pada Tabel 2.

Tabel 2. Hasil case folding, cleaning, dan tokenizing

| No. | Hasil CaseFolding | Hasil Cleaning | Hasil Tokenizing |
|------|--|--|--|
| 1 | @adhiweecak selamat pagi kak. penumpang ka usia 17 tahun atau lebih saat boarding wajib menunjukkan identitas aktif asli dan berfoto digital dan wajib dari aplikasi dukcapil (dinas kependudukan dan pencatatan sipil) yang ter | selamat pagi kak penumpang ka usia tahun atau lebih saat boarding wajib menunjukkan identitas aktif asli dan berfoto digital dan wajib dari aplikasi dukcapil dinas kependudukan dan pencatatan sipil yang ter | ['selamat', 'pagi', 'kak', 'penumpang', 'kakak', 'usia', 'tahun', 'atau', 'lebih', 'saat', 'boarding', 'wajib', 'menunjukkan', 'identitas', 'aktif', 'asli', 'dan', 'berfoto', 'silakan', 'dapat', 'menunjukkan', 'digital', 'dan', 'wajib', 'dari', 'aplikasi', 'dukcapil', 'dinas', 'kependudukan', 'dan', 'pencatatan', 'sipil', 'yang', 'tersedia', 'di', 'perangkat', 'handphone', 'milik'] |
| 2 | @kai121 apakah bisa check in kaji dengan menggunakan ktp yg tersimpan dalam aplikasi identitas kependudukan digital yang mana aplikasi tersebut adalah aplikasi resmi dari kementerian dalam negeri? | apakah bisa check in kaji dengan menggunakan ktp yg tersimpan dalam aplikasi identitas kependudukan digital yang mana aplikasi tersebut adalah aplikasi resmi dari kementerian dalam negeri | ['apakah', 'bisa', 'cek', 'ini', 'kai', 'dengan', 'menggunakan', 'yang', 'tersimpan', 'dalam', 'aplikasi', 'identitas', 'kependudukan', 'digital', 'yang', 'mana', 'aplikasi', 'tersebut', 'adalah', 'aplikasi', 'resmi', 'dari', 'kementerian', 'dalam', 'negeri'] |
| ... | ... | ... | ... |
| 6696 | @kompascom nah bagus ini. kedepanya tidak ada lagi namanya fotokopi ktp. beneran primitif gais kalo kita apa masih harus fotokopi. di negara maju semua serba digital m cardless. satu data bisa diakses untuk kebutuhan apapun. #ktpdigital | nah bagus ini kedepanya tidak ada lagi namanya fotokopi ktp beneran primitif gais kalo kita apa masih harus fotokopi di negara maju semua serba digital m cardless satu data bisa diakses untuk kebutuhan apapun | ['nah', 'bagus', 'ini', 'kedepanya', 'tidak', 'ada', 'lagi', 'namanya', 'fotokopi', 'beneran', 'primitif', 'gais', 'kalo', 'kita', 'apa', 'masih', 'harus', 'fotokopi', 'di', 'negara', 'maju', 'semua', 'serba', 'digital', 'sama', 'cardless', 'satu', 'data', 'bisa', 'diakses', 'untuk', 'kebutuhan', 'apapun'] |
| 6697 | kota jambi akan terapkan ktp digital kadis dukcapil: tak perlu cetak via https://t.co/jzaixlgnyf via @tribunjambiku #ktp #ktpdigital #kotajambi #dukcapil | kota jambi akan terapkan ktp digital kadis dukcapil tak perlu cetak via | ['kota', 'jambi', 'akan', 'terapkan', 'digital', 'dukcapil', 'tidak', 'perlu', 'cetak', 'melalui'] |

d. Normalisasi

Normalisasi dilakukan untuk mengubah kata-kata yang tidak baku menjadi kata baku dengan menggunakan 5 jenis kamus. Hasil ditunjukkan pada Tabel 3.

e. Stopword removal dilakukan untuk menghapus kata-kata seperti "yang", "atau", "dan" yang tidak mempengaruhi arti pada sentimen dan pemahaman pada suatu kalimat. Selanjutnya dilakukan tahap stemming untuk mengurangi jumlah kata kunci yang berbeda dalam sebuah dokumen. Hasil ditunjukkan pada Tabel 3.

f. Stemming dilakukan untuk menghilangkan kata imbuhan menjadi bentuk dasar. Hasil ditunjukkan pada Tabel 3.

Tabel 3. Hasil normalisasi, stopwrd removal, dan stemming

| No. | Hasil Normalisasi | Hasil Stopword Removal | Hasil Stemming |
|-----|---|--|---|
| 1 | selamat pagi kak penumpang kakak usia tahun atau lebih saat boarding wajib menunjukkan identitas aktif asli dan berfoto silakan dapat menunjukkan digital dan wajib dari aplikasi dukcapil dinas kependudukan dan pencatatan sipil yang | ['selamat', 'pagi', 'kak', 'penumpang', 'kakak', 'usia', 'boarding', 'wajib', 'identitas', 'aktif', 'asli', 'berfoto', 'silakan', 'digital', 'wajib', 'aplikasi', 'dukcapil', 'dinas', 'kependudukan', 'pencatatan', 'sipil', 'tersedia', 'perangkat', 'handphone', 'milik'] | selamat pagi kak tumpang kakak usia boarding wajib identitas aktif asli foto sila digital wajib aplikasi dukcapil dinas duduk catat sipil sedia perangkat handphone milik |

| No. | Hasil Normalisasi | Hasil Stopword Removal | Hasil Stemming |
|------|---|--|---|
| 2 | tersedia di perangkat handphone milik apakah bisa cek ini kai dengan menggunakan yang tersimpan dalam aplikasi identitas kependudukan digital yang mana aplikasi tersebut adalah aplikasi resmi dari kementerian dalam negeri | ['cek', 'kai', 'tersimpan', 'aplikasi', 'identitas', 'kependudukan', 'digital', 'aplikasi', 'aplikasi', 'resmi', 'kementerian', 'negeri'] | cek kai simpan aplikasi identitas duduk digital aplikasi aplikasi resmi menteri negeri |
| 6696 | nah bagus ini kedepanya tidak ada lagi namanya fotokopi benaran primitif gais kalau kita apa masih harus fotokopi di negara maju semua serba digital sama cardless satu data bisa diakses untuk kebutuhan apapun | ['bagus', 'kedepanya', 'namanya', 'fotokopi', 'benaran', 'primitif', 'gais', 'fotokopi', 'negara', 'maju', 'serba', 'digital', 'cardless', 'data', 'diakses', 'kebutuhan', 'apapun'] | bagus depa nama fotokopi benar primitif gais fotokopi negara maju serba digital cardless data akses butuh apa |
| 6697 | kota jambi akan terapkan digital dukcapil tidak perlu cetak melalui | ['kota', 'jambi', 'terapkan', 'digital', 'dukcapil', 'cetak'] | kota jambi terap digital dukcapil cetak |

3.2 Pelabelan Data

Pelabelan data dilakukan untuk memberi label pada data yang telah melalui proses preprocessing. Proses pelabelan data membutuhkan lexicon-based. Kamus yang digunakan dalam penelitian analisis sentimen ini adalah SenticNet7. Hasil pelabelan data positif, negatif, dan netral ditunjukkan pada Tabel 4.

Tabel 4. Hasil Pelabelan Data

| Positif | Negatif | Netral |
|---------|---------|--------|
| 3625 | 2726 | 346 |

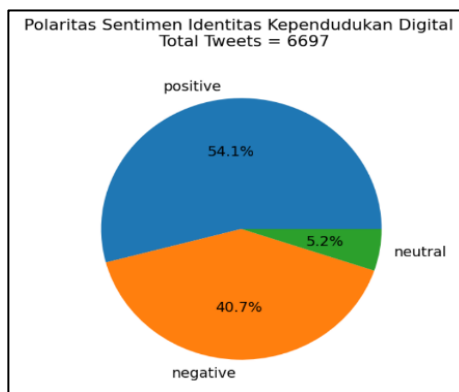
Kemudian, untuk hasil penghapusan data netral ditunjukkan pada Tabel 5.

Tabel 5. Hasil Pelabelan Data tanpa Data Netral

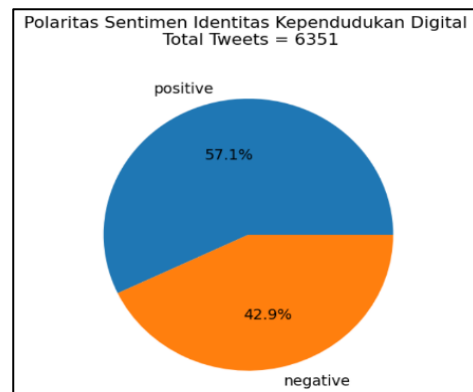
| Positif | Negatif |
|---------|---------|
| 3265 | 2726 |

3.3 Visualisasi Data

Visualisasi data menampilkan hasil dari piechart dan wordcloud dalam memvisualisasikan hasil analisis klasifikasi, dan membandingkan frekuensi kata sentimen positif dan negatif. Hasil diagram pie untuk pelabelan data positive, negative, dan netral ditunjukkan pada Gambar 2a dan 2b.



Gambar 2a. Persentase Sentimen IKD



Gambar 2b. Persentase Sentimen IKD tanpa Sentimen Netral

Gambar 2a menunjukkan sentimen yang diperoleh pada tahap pelabelan dengan nilai positive sebesar 54,1% atau 3265 data opini, nilai negative sebesar 40,7% atau 2726 data opini, dan nilai netral sebesar 5,2% atau

346 data opini. Langkah berikutnya, yaitu penghapusan data netral. Penelitian ini hanya dilakukan pada sentimen positive dan negative. Sehingga, sentimen dengan polaritas netral dihapus. Ditunjukkan pada Gambar 2b.

Gambar 3b menunjukkan hasil digram pie dengan nilai positive sebesar 57,1% dan negative sebesar 42,9%. Selanjutnya, menampilkan sentimen dalam bentuk wordcloud. Data yang ditampilkan merupakan representasi visual dari frekuensi kemunculan kata pada suatu kata yang sering muncul. Besar kecilnya huruf menentukan frekuensi kemunculan suatu kata, sehingga semakin besar ukuran huruf, semakin besar kemunculan kata tersebut. Hasil wordcloud untuk data sentimen positive dan negative ditunjukkan pada Gambar 3.

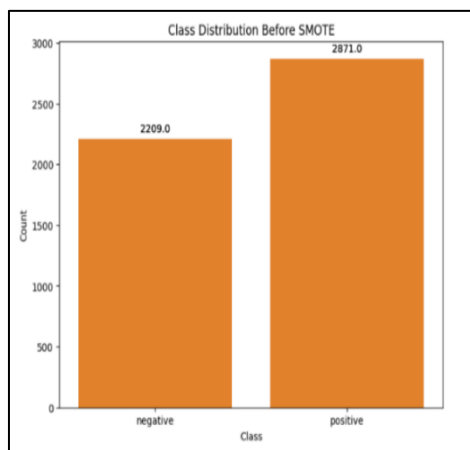


Gambar 3. Hasil Wordcloud Data Sentimen Positif dan Negatif

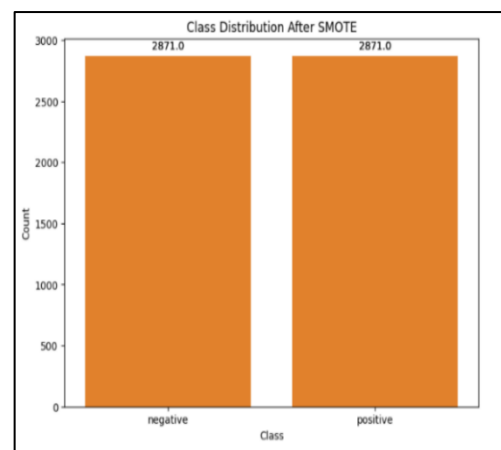
Gambar 3 menunjukkan hasil wordcloud untuk data sentimen positive dan negative. Kata-kata yang muncul pada wordcloud data sentimen positif, yaitu “ terima kasih”, “aktivasi”, “ikd”, “pakai”, “kartu digital”, “data”. Berikut adalah contoh tweet data sentimen positif, “**Terima kasih** @jabarprovgoind atas bantuan @disdukcapildepok Mobil Pelayanan Keliling yang berguna untuk memudahkan masyarakat dalam melakukan perekaman E-KTP serta aktivasi Identitas Kependudukan Digital”, “Tidak perlu khawatir E-KTP hilang atau rusak **pakai Identitas Kependudukan Digital (IKD)** formatnya sudah digital di hp lho! Yuk aktifin segera”, “Cara **aktivasi** sangat mudah lho cukup download Identitas Kependudukan Digital di playstore isi data diri sesuai KTP-El dan memiliki email aktif di Hp kemudian scan barcode untuk **aktivasi** ke petugas **IKD** yang berada di Kantor Dukcapil di Kecamatan Setempat. Sedangkan kata-kata yang muncul pada wordcloud data sentimen negatif, yaitu “bawa”, “ganti kartu”, “kena biaya”, “nomor handphone”, “bawa dokumen”, dan lain-lain. Berikut adalah contoh data sentimen negatif, “**Bawa** KTP digital kira petugasnya paham / tidak ? terima atau tidak? Ingat ! Digitalisasi tidak hanya konsep saja. Tapi diimplementasikan. SDM juga harus pintar (smart). KTP cetakan sudah disimpan di rumah. Yang digital ada di aplikasi Identitas Kependudukan Digital (IKD)”.

3.4 Optimasi SVM Menggunakan SMOTE

Penelitian ini menghasilkan data sentimen positive dan negative yang tidak seimbang. Oleh karena itu, dilakukan SMOTE untuk menghasilkan data sentimen yang seimbang. Hasil data sentimen sebelum dan sesudah menggunakan SMOTE ditunjukkan pada Gambar 4a dan 4b.



Gambar 4a. Hasil Optimasi SVM Sebelum Menggunakan SMOTE



Gambar 4b. Hasil Optimasi SVM Setelah Menggunakan SMOTE

Gambar 5a menampilkan hasil optimasi SVM sebelum menggunakan SMOTE dengan data sentimen positif sebanyak 2871 dan negatif sebanyak 2209 data. Langkah selanjutnya, menyeimbangkan data sentimen positive dan negative yang ditunjukkan pada Gambar 7. Sedangkan untuk Gambar 5b menampilkan hasil optimasi SVM setelah menggunakan SMOTE dengan data sentimen positif dan negatif menghasilkan nilai yang sama yaitu

sebanyak 2871 data. Data sentimen yang diperlukan untuk menghasilkan nilai yang seimbang yaitu sebanyak 662 data.

3.5 Pemodelan Support Vector Machine (SVM)

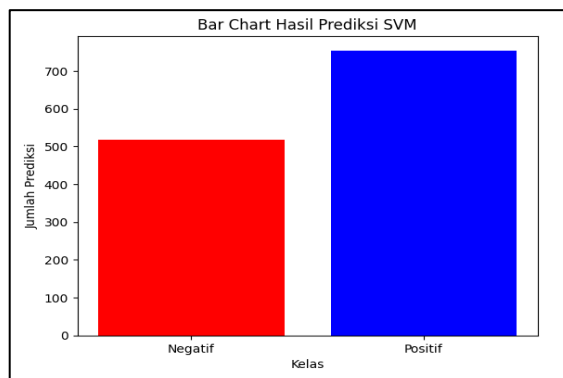
Tahap Pemodelan dilakukan untuk menghitung klasifikasi pada penelitian analisis sentimen ini dengan pembuatan data uji dan data latih. Ditunjukkan pada Tabel 6.

Tabel 6. Hasil Pemodelan Support Vector Machine (SVM)

| | Data Uji | Data Latih |
|------------|----------|------------|
| Presentase | 20% | 80% |
| Jumlah | 1148 | 4594 |

3.6 Klasifikasi Menggunakan Support Vector (SVM)

Klasifikasi menggunakan SVM digunakan setelah menentukan data uji dan data latih. Bar chart hasil prediksi SVM ditunjukkan pada Gambar 5.

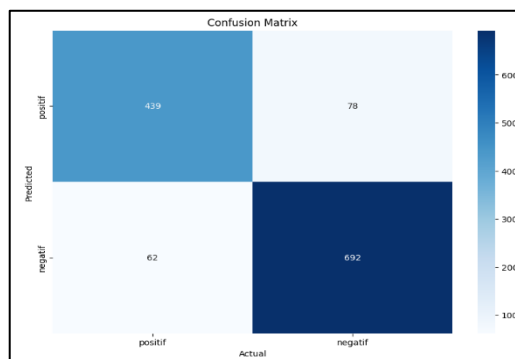


Gambar 5. Barchart hasil prediksi SVM

Gambar 5 menunjukkan barchart hasil prediksi SVM untuk membandingkan antara kelas positif dan negatif dalam perhitungan klasifikasi SVM. Model dapat mengklasifikasikan opini positif ke dalam klasifikasi sentimen positif sebanyak 754 opini. Sedangkan komentar negatif 517 opini.

3.7 Evaluasi

Tahap Evaluasi dilakukan menggunakan metode confusion matrix berdasarkan data uji yang akan diproses oleh model klasifikasi SVM. Selanjutnya, dilakukan perhitungan accuracy, presission, dan recall. Hasil dari evaluasi ditunjukkan pada Gambar 6 dan 7.



Gambar 6. Hasil Evaluasi Confusion Matrix

| Laporan Klasifikasi: | | | | |
|----------------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| negatif | 0.88 | 0.85 | 0.86 | 517 |
| positif | 0.90 | 0.92 | 0.91 | 754 |
| accuracy | | | 0.89 | 1271 |
| macro avg | 0.89 | 0.88 | 0.89 | 1271 |
| weighted avg | 0.89 | 0.89 | 0.89 | 1271 |

Gambar 7. Hasil Evaluasi Laporan Klasifikasi



Gambar 6 merujuk hasil evaluasi confusion matrix dengan nilai TP = 439, TN = 692, FP = 62, FN = 78 maka disimpulkan $439 + 692 = 1131$ data diklasifikasikan dengan benar dan $62 + 78 = 140$ data diklasifikasikan dengan tidak benar dari total 1271 data dan pada Gambar 7 terdapat hasil evaluasi laporan klasifikasi dengan accuracy sebesar 89%, presisi negatif sebesar 88%, presisi positif sebesar 90%, recall negatif sebesar 85%, recall positif sebesar 92%, f1-score negatif sebesar 86%, dan f1-score positif sebesar 91%. Dengan demikian, disimpulkan bahwa SVM yang dibangun mendapatkan hasil keefektifan yang sangat baik dalam mengklasifikasikan sentimen terhadap IKD.

4. KESIMPULAN

Analisis sentimen terhadap IKD pada media sosial X memperoleh data opini masyarakat dengan teknik crawling menggunakan bahasa pemrograman python pada google colab. Data opini diambil dari tahun 2021 sampai tahun 2024 sebanyak 6697 opini masyarakat. Dalam melakukan preprocessing akan melewati beberapa tahapan, yaitu case folding, cleaning, tokenizing, normalisasi, stopword removal, dan stemming. Setelah melakukan preprocessing dilanjut dengan pelabelan data, visualisasi data, optimasi SVM menggunakan SMOTE, klasifikasi menggunakan SVM, dan yang terakhir adalah evaluasi. Berdasarkan hasil klasifikasi 1.148 data uji menggunakan SVM didapatkan sentimen positive sebanyak 754 opini, sedangkan klasifikasi sentimen negative sebanyak 517 opini. Selain itu, model yang dikembangkan dapat melakukan klasifikasi secara baik. Hal tersebut ditunjukkan pada nilai accuracy sebesar 89%, recall positif sebesar 92% serta recall negatif sebesar 85%. Sementara itu, kemampuan model dalam melakukan prediksi sangat baik dengan nilai precision sebesar 89%. Saran penelitian selanjutnya adalah perbaikan pada aspek normalisasi kalimat. Pre-trained word embedding seperti IndoBert dapat digunakan untuk proses normalisasi.

REFERENCES

- [1] Peraturan Menteri Dalam Negeri, “Standar Dan Spesifikasi Perangkat Keras, Perangkat Lunak, Dan Blangko Kartu Tanda Penduduk Elektronik Serta Penyelenggaraan Identitas Kependudukan Digital” 2022
- [2] I. B. Permadi and A. Rokhman, “Implementasi Identitas Kependudukan Digital Dalam Upaya Pengamanan Data pribadi,” JOPPAS: Journal of Public Policy and Administration Silampari, vol. 4, no. 2, pp. 80–88, Jun. 2023, doi: 10.31539/joppas.v4i2.6199.
- [3] B. Putri and O. Reviandani, “Penerapan E-Government Melalui Pelaksanaan Program Ktp Digital Di Kelurahan Dr. Soetomo Kota Surabaya,” 2023. [Online]. Available: <http://journal.uta45jakarta.ac.id/index.php/admpublik/ix>
- [4] C. Mole, E. Chalstrey, P. Foster, and T. Hobson, “Digital identity architectures: comparing goals and vulnerabilities,” Feb. 2023, [Online]. Available: <http://arxiv.org/abs/2302.09988>
- [5] M. Alfarizi, “Digitalisasi Kartu Tanda Penduduk dan Partisipasi Milenial-Gen Z: Investigasi Penerimaan Transformasi Digital dalam Kebijakan Kependudukan Indonesia,” Jurnal Studi Kebijakan Publik, vol. 2, no. 1, pp. 41–54, May 2023, doi: 10.21787/jskp.2.2023.41-54.
- [6] M. Riyanto and V. Kovalenko, “Partisipasi Masyarakat Menuju Negara Kesejahteraan: Memahami Pentingnya Peran Aktif Masyarakat Dalam Mewujudkan Kesejahteraan Bersama,” 2023.
- [7] D. Fitriani et al., “Partisipasi Masyarakat Dalam Proses Demokrasi Di Indonesia: Analisis Peran Teknologi Dan Media Sosial,” 2023.
- [8] D. R. Sari, B. Matsaany, and M. Hamka, “Aspect Extraction Of E-Commerce And Marketplace Applications Using Word2vec And Wordnet Path,” Jurnal Teknik Informatika (Jutif), vol. 4, no. 4, pp. 787–796, Aug. 2023, doi: 10.52436/1.jutif.2023.4.4.726.
- [9] 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.
- [10] R. A. Lestari, A. Erfina, W. Jatmiko, and P. Korespondensi, “Penerapan Algoritma Support Vector Machine Pada Analisis Sentimen Terhadap Identitas Kependudukan Digital,” vol. 10, no. 5, pp. 1063–1070, 2023, doi: 10.25126/jtiik.2023107264.
- [11] R. Hidayat, R. Nur Rahman, M. Reifin Perdana, P. Teknik Informatika, F. Sains dan Teknologi, and U. Muhammadiyah Kalimantan Timur, “Analisis Sentimen Aplikasi Identitas Kependudukan Digital (IKD) Menggunakan Metode Naïve Bayes,” Jurnal Sistem Informasi dan Ilmu Komputer, vol. 2, no. 1, pp. 129–140, 2024, doi: 10.59581/jusiik-widyakarya.v2i1.2320.
- [12] V. Fitriyana et al., “Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine,” 2023.
- [13] N. Fitriyah, B. Warsito, D. Asih, and I. Maruddani, “Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm),” Jurnal Gaussian, vol. 9, no. 3, pp. 376–390, 2020, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/>
- [14] B. Ramadhani, R. R. Suryono, and K. Kunci, “Jurnal Media Informatika Budidarma Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse,” 2024, doi: 10.30865/mib.v8i2.7458.
- [15] S. Aftab, S. Shah Muhammad, S. Awan, M. Ahmad, and S. Ahmad, “Machine Learning Techniques for Sentiment Analysis: A Review,” International Journal Of Multidisciplinary Sciences And Engineering, vol. 8, no. 3, 2017, [Online]. Available: www.ijmse.org



- [16] R. Pappagari, P. Zelasko, J. Villalba, Y. Carmiel, and N. Dehak, “Hierarchical Transformers for Long Document Classification,” in 2019 IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2019 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Dec. 2019, pp. 838–844. doi: 10.1109/ASRU46091.2019.9003958.
- [17] M. Ahmad, S. Aftab, M. S. Bashir, N. Hameed, I. Ali, and Z. Nawaz, “SVM optimization for sentiment analysis,” *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 4, pp. 393–398, 2018, doi: 10.14569/IJACSA.2018.090455.
- [18] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, “Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang,” *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 677–690, Jul. 2022, doi: 10.30812/matrik.v21i3.1726.
- [19] A. Surya Firmansyah, A. Aziz, and M. Ahsan, “Optimasi K-Nearest Neighbor Menggunakan Algoritma Smote Untuk Mengatasi Imbalance Class Pada Klasifikasi Analisis Sentimen,” 2023.
- [20] Y. Xu and R. Xu, “Research on Interpolation and Data Fitting: Basis and Applications,” Aug. 2022, [Online]. Available: <http://arxiv.org/abs/2208.11825>
- [21] D. Aby Vonega, A. Fadila, and D. Ely Kurniawan, “Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani Dalam Pilpres 2024,” 2022. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [22] E. Febriyani and H. Februriyanti, “Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter,” vol. 17, no. 1, 2023
- [23] “Titian: Jurnal Ilmu Humaniora,” 2023, [Online]. Available: <https://online-journal.unja.ac.id/index.php/titian>
- [24] D. Haryani, “Kualitas Pelayanan Kartu Tanda Penduduk Elektronik (Ktp-El) Di Kecamatan Banyuasin Kabupaten Banyuasin,” 2024.
- [25] M. I. Amal, E. S. Rahmasita, E. Suryaputra, and N. A. Rakhmawati, “Analisis Klasifikasi Sentimen Terhadap Isu Kebocoran Data Kartu Identitas Ponsel di Twitter,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 3, Dec. 2022, doi: 10.28932/jutisi.v8i3.5483.
- [26] W. Ayunda Sari, D. Arifianto, and A. Nilogiri, “Analisis Sentimen Terhadap Ulasan Netizen Pada Aplikasi CamScanner Menggunakan Metode Support Vector Machine (SVM).” 2023
- [27] R. Tineges, A. Triayudi, and I. D. Sholihati, “Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM),” *Jurnal Media Informatika Budidarma*, Vol. 4, No. 3, P. 650, Jul. 2020, Doi: 10.30865/mib.v4i3.2181.
- [28] S. A. Ashari, M. W. A. Saputra, E. Larosa, and B. S. Rijal, “Analisis Sentimen pada Aplikasi Translate Google Menggunakan Metode SVM (Studi Kasus: Komentar Pada Playstore),” *Jurnal Teknik*, vol. 21, no. 2, pp. 168–182, Dec. 2023, doi: 10.37031/jt.v21i2.412.
- [29] F. Zuhad and N. Wilantika, “Perbandingan Penggunaan Kamus Normalisasi dalam Analisis Sentimen Berbahasa Indonesia.” 2022
- [30] S. Fachri and P. J. Ramdan, “Pemodelan Machine Learning : Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Menggunakan Data Twitter,” 2022. [Online]. Available: <https://t.co/IEnuGFuuJ>
- [31] D. R. Sari, B. Matsaany, and M. Hamka, “Aspect Extraction Of E-Commerce And Marketplace Applications Using Word2vec And Wordnet Path,” *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 4, pp. 787–796, Aug. 2023, doi: 10.52436/1.jutif.2023.4.4.726.
- [32] M. Hamka and D. Ratna Sari, “Analisis Sentimen Dan Information Extraction Pembelajaran Daring Menggunakan Pendekatan Lexicon,” 2022.
- [33] A. Firdaus, “Analisis Sentimen Pada Aplikasi Alfagift.” UIN Jakarta, 2023
- [34] R. Parluka, S. Ilham Pradika, A. M. Hakim, and R. N. M. Kholilul, “Analisis Sentimen Twitter Terhadap Bitcoin dan Cryptocurrency Berbasis Python TextBlob,” 2020. [Online]. Available: <https://t.co/QaUW3P2TKc>
- [35] S. Abdullah, “Penulis Pertama: Visualisasi Data Analisa Sentimen ... 261 Visualisasi Data Analisa Sentimen RUU Omnibus law Kesehatan Menggunakan KNN dengan Software RapidMiner,” vol. 8, no. 3, 2023.
- [36] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiyari, “Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE,” *AITI: Jurnal Teknologi Informasi*, vol. 18, no. Agustus, pp. 173–184, 2021.
- [37] R. Turaina, R. Saputra, S. Informasi, U. Metamedia, and J. Khatib Sulaiman Dalam, “Optimalisasi Klasifikasi Umpan Balik Mahasiswa Terhadap Layanan Kampus dengan Sinergi Random Forest dan Smote,” *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 6, no. 6, 2023.
- [38] D. Hana Amalia and W. Yustanti, “Klasifikasi Buku Menggunakan Metode Support Vector Machine pada Digital Library,” *Journal of Informatics and Computer Science*, vol. 03, 2021
- [39] J. Teknika, M. K. Rifa, M. H. Totohendarto, and M. R. Muttaqin, “Analisis Sentimen Pengguna E-Wallet Dana Dan Gopay Pada Twitter Menggunakan Metode Support Vector Machine (SVM),” *IJCCS*, 2023
- [40] F. Setya Ananto and F. N. Hasan, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store,” *Jurnal ICT : Information Communication & Technology*, vol. 23, no. 1, pp. 75–80, 2023, [Online]. Available: <https://ejournal.ikmi.ac.id/index.php/jict-ikmi>