

PERBANDINGAN METODE KLASIFIKASI DALAM ANALISIS SENTIMEN MASYARAKAT TERHADAP IDENTITAS KEPENDUDUKAN DIGITAL (IKD)

Novia Wahyuningsih*¹⁾, Hendry²⁾

1. Universitas Kristen Satya Wacana, Salatiga, Indonesia
2. Universitas Kristen Satya Wacana, Salatiga, Indonesia

Article Info

Kata Kunci: *Confusion Matrix*, CRISP-DM, Klasifikasi, *Orange Data Mining*, Sentimen

Keywords: *Confusion Matrix*, CRISP-DM, Classification, *Orange Data Mining*, Sentiment

Article history:

Received 12 May 2023

Revised 26 May 2023

Accepted 9 June 2023

Available online 1 December 2023

DOI :

<https://doi.org/10.29100/jipi.v8i4.4155>

* Corresponding author.

Novia Wahyuningsih

E-mail address:

672019296@student.uksw.edu

ABSTRAK

Identitas Kependudukan Digital merujuk pada penggunaan teknologi digital dan data elektronik untuk mengidentifikasi individu dalam konteks kependudukan. Pada dasarnya, Identitas Kependudukan Digital bertujuan untuk memberikan cara yang lebih efisien dalam mengelola data identitas individu. Dalam mengimplementasikan Identitas Kependudukan Digital, perlu memperhatikan aspek keamanan dan privasi data. Selain itu, tidak semua individu memiliki akses ke teknologi digital atau mungkin menghadapi tantangan dalam menggunakan teknologi tersebut. Penelitian ini bertujuan untuk melakukan klasifikasi data dari Twitter terkait Identitas Kependudukan Digital dengan membandingkan metode SVM, K-NN, *Naive Bayes*, dan *Neural Network* menggunakan pendekatan *data mining* CRISP-DM. Dataset diambil menggunakan Twitter API dan diproses menggunakan *Orange Data Mining*. Dari jumlah awal data *tweet* sebanyak 7914, setelah dilakukan pembersihan data, tersisa 1022 *tweet* yang digunakan dalam penelitian. Hasil pengujian menunjukkan bahwa masyarakat cenderung memiliki sentimen netral terkait Identitas Kependudukan Digital. Dalam hal performa model klasifikasi, metode K-NN menunjukkan kinerja yang sangat baik dengan nilai akurasi, presisi, dan *recall* mencapai 100%. Metode *Neural Network* dan *Naive Bayes* memiliki perbedaan yang kecil dalam performanya, sementara metode SVM memiliki nilai yang lebih rendah dalam evaluasi menggunakan *Confusion Matrix*. Penelitian ini memberikan wawasan tentang sentimen masyarakat terkait Identitas Kependudukan Digital dan mengidentifikasi metode klasifikasi dengan performa terbaik. Hasilnya dapat digunakan sebagai dasar untuk pengembangan lebih lanjut dalam meningkatkan akses dan pengelolaan Identitas Kependudukan Digital.

ABSTRACT

Digital Population Identity refers to using digital technology and electronic data to identify individuals in a residential context. Fundamentally, Digital Population Identity aims to provide a more efficient way of managing individual identity data. In implementing Digital Population Identity, it is necessary to pay attention to aspects of data security and privacy. In addition, not all individuals have access to digital technologies or may face challenges in using these technologies. This research aims to classify data from Twitter related to Digital Population Identity by comparing SVM, K-NN, Naive Bayes, and Neural Network methods using the CRISP-DM data mining approach. Datasets are retrieved using the Twitter API and processed using Orange Data Mining. From the initial number of tweets data of 7914, the remaining 1022 tweets were used in the study after data cleaning. The test results show that people tend to have neutral sentiments regarding Digital Population Identity. In terms of classification model performance, the K-NN method exhibits excellent performance with accuracy, precision and recall reaching 100%. The Neural Network and Naive Bayes methods have little difference in their performance, while the SVM method has a lower value in the evaluation using the Confusion Matrix. This research provides insight into public sentiment regarding Digital Population Identity and identifies the best performing classification methods. The results can be used as a basis for further development in increasing access to and managing digital identities.

I. PENDAHULUAN

SETIAP individu yang merupakan warga negara Indonesia memiliki kesempatan untuk memperoleh layanan administrasi kependudukan. Sejalan dengan kemajuan teknologi, pemerintah menyediakan pelayanan yang cepat dan efisien kepada seluruh warga negara, yang pada gilirannya meningkatkan pengelolaan data. Untuk mempercepat proses pendokumentasian kependudukan, pemerintah menciptakan inovasi baru yang bertujuan meningkatkan pengurusan dokumen setiap tahunnya. Salah satu inovasi yang dikembangkan untuk digitalisasi dokumen kependudukan adalah Identitas Kependudukan Digital (IKD).

Identitas Kependudukan Digital (IKD) merupakan data kependudukan berbasis digital yang dikenalkan dalam bentuk aplikasi perangkat seluler berupa foto, ataupun *QR Code* [1]. Pemerintah berharap dengan adanya aplikasi Identitas Kependudukan Digital (IKD), dokumen kependudukan tidak perlu lagi dicetak atau disimpan dalam bentuk berkas fisik Kartu Tanda Penduduk (KTP) untuk keperluan administrasi dan juga dapat menghemat anggaran penggunaan blangko KTP-el [2]. Aplikasi ini terhubung dan terintegrasi dengan layanan kesehatan, pendidikan, layanan sosial, perbankan, pajak, *payment gateway*, dan sebagainya dalam aktivasi tunggal yang terdapat dalam fitur IKD, sehingga mudah digunakan oleh masyarakat di era perkembangan teknologi [2].

Meskipun inovasi ini diharapkan dapat meningkatkan efisiensi dan kualitas pengelolaan data kependudukan, terdapat perdebatan di masyarakat terkait implementasi dan keamanan Identitas Kependudukan Digital (IKD). Beberapa masyarakat, terutama di daerah pelosok, merasa khawatir karena keterbatasan akses internet dan minimnya sosialisasi mengenai aplikasi ini. Keraguan juga timbul terkait keamanan data identitas dalam bentuk digital. Dalam konteks ini, media sosial, khususnya Twitter, menjadi saluran yang digunakan masyarakat untuk menyampaikan pendapat, opini, dan komentar terkait Identitas Kependudukan Digital (IKD).

Penelitian terdahulu yang menjadi rujukan dari penelitian ini menganalisis opini masyarakat tentang kerjasama antara restoran cepat saji McDonald's dengan *boy band* terkenal BTS, yang menghasilkan fenomena antrian panjang di depan setiap cabang McDonald's oleh para pengemudi ojek *online*. Penelitian tersebut menggunakan pendekatan analisis distribusi *clustering* dan *Orange Data Mining* sebagai alat untuk memvisualisasikan sentimen dalam setiap *corpus* data Twitter. Hasil penelitian tersebut menyimpulkan bahwa analisis *clustering* dapat digunakan untuk mengklasifikasikan pengguna Twitter berdasarkan sentimen mereka [3]. Dari penjabaran penelitian tersebut, maka penelitian ini mengadopsi pendekatan menggunakan *Orange Data Mining* sebagai *tools* untuk melakukan *preprocessing* dan klasifikasi data *tweet*.

Dalam sebuah penelitian sebelumnya membandingkan tiga metode klasifikasi, yaitu *Neural Network*, *Support Vector Machine*, dan *Naive Bayes* dalam mendeteksi kanker payudara menggunakan *dataset Wisconsin Breast Cancer*. Hasil penelitian menunjukkan bahwa menggunakan *Support Vector Machine* (SVM) memiliki tingkat akurasi klasifikasi yang lebih tinggi daripada dua algoritma lainnya. Akurasi yang dicapai oleh SVM mencapai 98,9%. Oleh karena itu, penelitian ini menyimpulkan bahwa SVM merupakan model terbaik untuk prediksi kanker payudara jika dibandingkan dengan *Neural Network* dan *Naive Bayes* [4]. Dari hasil penelitian tersebut menjadi pertimbangan bagi penulis dalam melakukan penelitian dengan menggunakan perbandingan metode klasifikasi lainnya.

Pada penelitian sebelumnya yang menjadi rujukan melibatkan perbandingan antara metode SVM dan K-NN dalam mengklasifikasikan kelulusan mahasiswa pada suatu mata kuliah. Tujuan penelitian tersebut adalah untuk membandingkan nilai AUC (*Area Under Curve*), akurasi, error, dan nilai ambang kelulusan mahasiswa antara kedua metode tersebut. Data *train* yang digunakan dalam penelitian ini didasarkan pada penilaian nyata oleh dosen terhadap mahasiswa, dengan hasil lulus atau tidak lulus sebagai variabel target. Hasil penelitian menunjukkan bahwa metode K-NN memiliki nilai AUC dan akurasi yang lebih tinggi dibandingkan dengan metode SVM. Selain itu, metode K-NN juga memiliki nilai error yang lebih rendah, menandakan adanya sedikit kesalahan dalam perhitungan data, namun metode SVM menunjukkan kinerja yang lebih baik dalam menentukan ambang kelulusan mahasiswa. Hal ini disebabkan adanya beberapa data *noise* yang mempengaruhi performa metode K-NN [5]. Dari penelitian tersebut, penulis memutuskan untuk memasukkan metode K-NN sebagai perbandingan dalam penelitian terkait Identitas Kependudukan Digital (IKD) dari media sosial Twitter.

Berdasarkan latar belakang yang telah dijelaskan, penelitian dapat dilakukan untuk mencari tahu sentimen yang dihasilkan dari topik bahasan mengenai Identitas Kependudukan Digital (IKD) dari media sosial Twitter dengan membandingkan metode K-NN, *Naive Bayes*, SVM, dan *Neural Network*. Dengan melakukan penelitian ini, diharapkan dapat memperoleh pemahaman yang lebih baik mengenai pandangan masyarakat terhadap Identitas Kependudukan Digital (IKD), serta memberikan wawasan yang berharga bagi pemerintah dalam pengembangan inovasi administrasi kependudukan untuk mencapai identitas serbaguna yang efisien dan aman. Pada penelitian ini, data diambil berdasarkan data *tweet* menggunakan *Application Programming Interface* (API) Twitter dengan melakukan *request* ke *Twitter Development* dan memanfaatkan *Orange Data Mining* sebagai *tools* dari metode

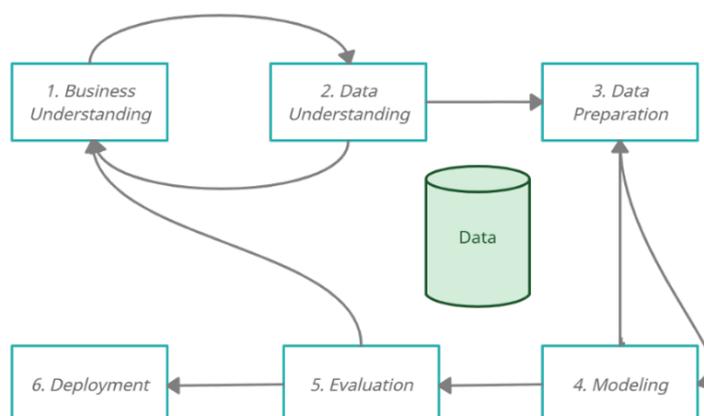
tersebut [6]. Data diambil dengan memasukkan *keywords* yang berhubungan dengan Identitas Kependudukan Digital (IKD) sehingga mendapatkan metode dengan tingkat performa paling baik. Dalam penelitian ini, *Orange Data Mining* digunakan sebagai *tools* analisis data karena fitur *preprocess text* yang dimilikinya. *Preprocess text* dalam *Orange Data Mining* difokuskan pada pengenalan tipe konten dari berbagai informasi *tweet* dengan menerapkan teknik klasifikasi dan pengelompokan (*clustering*) [7].

II. METODE PENELITIAN

Penelitian ini menggunakan metode perbandingan K-NN, *Naive Bayes*, SVM, dan *Neural Network* untuk menganalisis sentimen positif, negatif, dan netral berdasarkan data *tweet* yang terkait dengan Identitas Kependudukan Digital (IKD) di media sosial Twitter. Tujuan utama dari penelitian ini adalah untuk mengetahui sentimen yang muncul dalam topik Identitas Kependudukan Digital di *platform* Twitter. Dalam penelitian ini, peneliti menggunakan fakta dan pendapat pengguna Twitter saat ini sebagai sumber informasi untuk menjelaskan bagaimana masalah yang diteliti dapat dipecahkan. Objek penelitian yang dituju adalah konten *tweet* yang berhubungan dengan Identitas Kependudukan Digital (IKD). Untuk mengumpulkan data dari media sosial Twitter, peneliti menggunakan *tools* pengumpulan data yang disebut *Application Programming Interface* (API) Twitter, yang telah dipersiapkan sebelumnya. API ini berfungsi untuk mengakses dan mengenali pola komunikasi yang ada di Twitter [7].

Dalam analisis sentimen, peneliti menggunakan *Orange Data Mining* sebagai *tools* untuk melakukan analisis. *Orange Data Mining* adalah perangkat lunak yang digunakan untuk menganalisis dan memvisualisasikan data menggunakan teknik pembelajaran mesin yang bersifat *open source* [8]. Dengan menggunakan *Orange Data Mining*, peneliti dapat melihat kata-kata yang paling dominan muncul dalam komentar atau pendapat akun Twitter terkait, yang ditampilkan dalam bentuk *word cloud* [6]. Melalui penggunaan *Orange Data Mining*, peneliti dapat mengklasifikasikan dan mengidentifikasi kategori yang paling dominan dari sentimen positif, negatif, dan netral tanpa perlu melakukan pemrograman secara khusus (*coding*).

Dalam penelitian ini, digunakan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri dari beberapa tahapan. Tahapan-tahapan tersebut meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment* [9]. Setiap tahapan memiliki perannya masing-masing dalam proses penelitian, dan dapat digambarkan seperti yang terlihat pada Gambar 1, dengan penjelasan yang sesuai.



Gambar. 1. Metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*)

A. *Business Understanding* (Pemahaman Bisnis)

Pada tahap awal ini, peneliti mengklarifikasi tujuan dan batasan masalah dalam *data mining* dan melakukan perencanaan strategis. Penggunaan *data mining* dalam penelitian ini berkaitan dengan mengidentifikasi opini masyarakat di *platform* Twitter mengenai Identitas Kependudukan Digital (IKD).

B. *Data Understanding* (Pemahaman Data)

Pada tahap ini, dimulai proses pengumpulan data awal dan deskripsi data yang akan digunakan dalam penelitian. Dalam penelitian ini, digunakan *dataset* berupa *tweet* yang terkait dengan topik Identitas Kependudukan Digital (IKD). Pengumpulan *tweet* dilakukan dengan memasukkan *keyword* yang relevan dengan topik yang akan diteliti, seperti Dukcapil, KTP Digital, dan Identitas Kependudukan Digital.

C. Data Preparation (Persiapan Data)

Pada tahap persiapan data, dilakukan pengolahan *dataset* yang akan digunakan dalam tahap pemodelan. Data disiapkan untuk menjalani serangkaian langkah *preprocessing*, dimana data mentah yang telah dikumpulkan sebelumnya diubah menjadi bentuk yang siap digunakan. Proses *preprocessing* bertujuan untuk menghilangkan data yang tidak relevan dalam penelitian. Tahap *preprocessing* terdiri dari beberapa langkah, termasuk *Transformation*, *Tokenization*, *Normalization*, dan *Filtering* yang dilakukan menggunakan *Orange Data Mining* [10].

D. Modelling (Pemodelan)

Dalam tahap pemodelan ini, menggunakan perangkat lunak *Orange Data Mining*. Setelah melakukan pengujian model, *tweet* diklasifikasikan menjadi kategori positif, negatif, atau netral dengan membandingkan metode K-NN, *Naive Bayes*, SVM, dan *Neural Network*. Selain itu, juga dilakukan evaluasi terhadap nilai *Test and Score* dari hasil klasifikasi dengan metode tersebut. Setelah mengetahui nilai *Test and Score*, akan ditampilkan distribusi nilai atribut yang relevan [6].

E. Evaluation (Evaluasi)

Setelah menyelesaikan tahap pemodelan, dilakukan validasi untuk menguji model yang diajukan dan melakukan evaluasi pada *dataset* model menggunakan *Confusion Matrix* di dalam perangkat lunak *Orange Data Mining*. Proses ini melibatkan evaluasi kinerja pola yang dihasilkan dengan mempertimbangkan tingkat keakuratan [11].

F. Deployment (Penyebaran)

Pada tahap ini, informasi dari hasil analisis perbandingan metode klasifikasi yang diperoleh pada tahap sebelumnya dibuat ke dalam penulisan laporan sederhana dan artikel jurnal tentang hasil penelitian.

III. HASIL DAN PEMBAHASAN

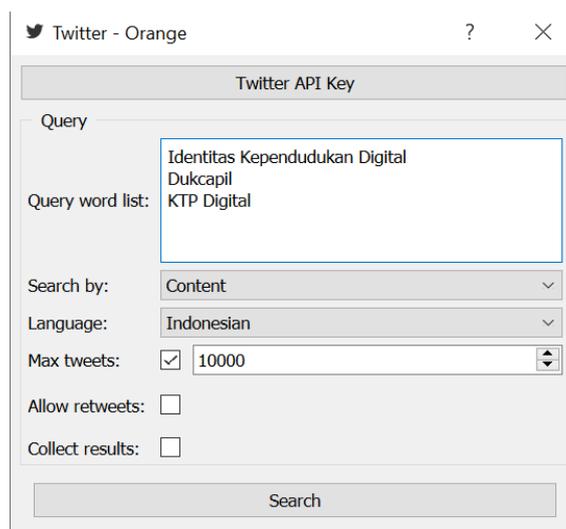
Pada penelitian yang dilakukan, bagian ini adalah tahap hasil dari uji analisis pada penelitian untuk membandingkan metode K-NN, *Naive Bayes*, SVM, dan *Neural Network* dalam mengklasifikasikan sentimen terkait topik Identitas Kependudukan Digital (IKD). Hasil dan pembahasan pada penelitian analisis sentimen yang telah dilakukan dijelaskan sebagai berikut :

A. Business Understanding

Data mining dalam penelitian ini diterapkan dengan melibatkan penggunaan data *tweet* yang terkait dengan Identitas Kependudukan Digital (IKD) untuk melakukan analisis sentimen dari positif, negatif maupun netral guna memahami opini dominan masyarakat terhadap IKD. Hasil analisis sentimen tersebut memberikan wawasan berharga bagi pemerintah dalam mengembangkan inovasi administrasi kependudukan yang efisien dan aman, dengan mempertimbangkan pandangan masyarakat terkait dengan IKD.

B. Data Understanding

Penelitian ini menggunakan *dataset* yang diambil dengan *Application Programming Interface* (API) Twitter. *Tools* yang dimanfaatkan untuk mendapatkan data *tweet* tersebut adalah *Orange Data Mining*. *Orange Data Mining* menggunakan *Design Widget* sebagai alur proses data [12]. Untuk mendapatkan data, maka diperlukan proses *Data Crawling Twitter* dengan memasukkan *keyword* tertentu yang berhubungan dengan topik yang akan diteliti. *Keyword* yang dimasukkan untuk mendapatkan data dalam penelitian ini adalah Identitas Kependudukan Digital, Dukcapil, KTP Digital. Dalam *Orange Data Mining*, data *input* yang digunakan disebut sebagai *corpus*. *Widget corpus* berfungsi sebagai kumpulan dokumen yang dapat menampilkan jumlah baris kalimat dan memungkinkan pengguna untuk menentukan fitur-fitur mana yang akan digunakan atau tidak digunakan dalam analisis [7]. Data *tweet* yang dihasilkan sebelum dilakukan *preprocessing* berjumlah 7914 *tweet*.



Gambar. 2. Data Text Mining Twitter

C. Data Preparation

Sebelum melakukan analisis teks, langkah awal yang dilakukan adalah *preprocess text*. Pada tahap ini, teks dipisahkan menjadi unit-unit yang lebih kecil yang disebut token, kemudian dilakukan *Transformation*, *Tokenization*, *Normalization*, dan *Filtering*. Urutan langkah-langkah ini dapat diaktifkan atau dinonaktifkan menggunakan *widget preprocess text* di *Orange Data Mining* [13]. *Preprocessing text* memastikan kualitas data yang baik sebelum tahap analisis dilakukan [14]. Adapun tahapan yang dilakukan dalam *preprocess text* dengan *Orange Data Mining* adalah sebagai berikut :

1) Transformation

Transformasi adalah proses mengubah data input menjadi bentuk yang diinginkan. Dalam konteks yang disebutkan, *transformasi* tersebut mengacu pada perubahan huruf secara *default*. Dengan demikian, dalam proses *transformasi*, setiap huruf pada data *input* akan diubah sesuai dengan tahapannya. Dalam *Transformation* terdapat beberapa tahap, antara lain :

- Lowercase* berfungsi untuk mengubah semua teks menjadi huruf kecil. Pada tahapan ini contohnya mengubah kata “Dukcapil” menjadi “dukcapil”.
- Remove Accents* untuk menghilangkan seluruh aksentuasi di dalam teks.
- Parse html* untuk mengetahui tag html dan mem-parsing teks.
- Remove url* untuk menghapus url dari teks.

2) Tokenization

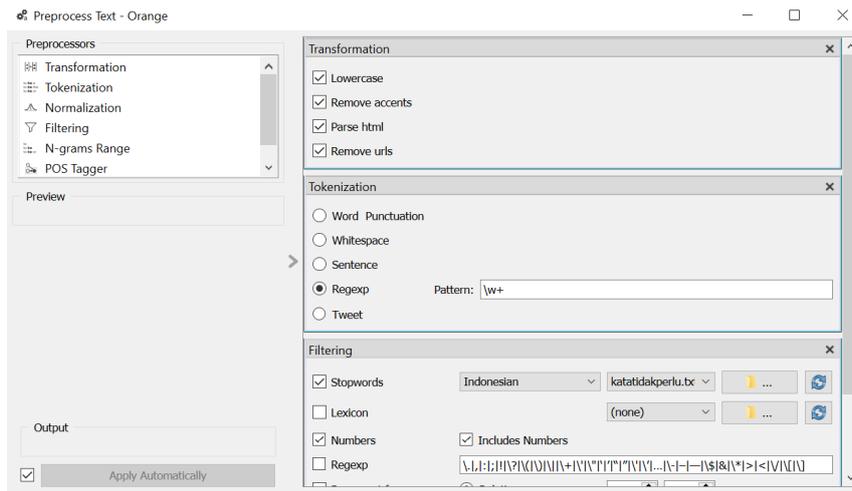
Tokenization merupakan metode memecah teks menjadi komponen yang lebih kecil. Biasanya proses ini dilakukan setelah tahapan *Transformation*. Tahapan ini akan dilakukan pembagian teks per kata dan mempertahankan simbol tanda baca. Artinya, teks akan dipecah menjadi kata-kata individu, dan simbol tanda baca seperti koma, titik, tanda tanya, dan sebagainya akan tetap dipertahankan dalam pemrosesan teks tersebut.

3) Normalization

Langkah ketiga dari *preprocess text* adalah *normalization* yang berlaku *stemming* dan *lemmatization text*. Teks yang telah dipisahkan kata per kata akan menjadi sebuah teks yang berdiri sendiri dalam suatu kalimat. Komentar atau pendapat biasanya mengandung penulisan yang kurang sempurna (*typo*). Pada tahap inilah mengenali penyusunan teks yang kurang sempurna tersebut.

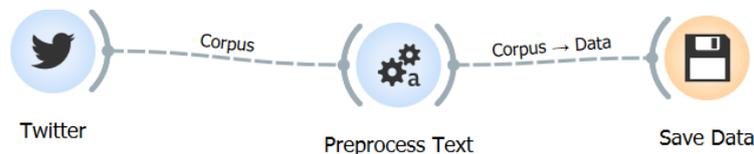
4) Filtering

Proses terakhir pada *preprocess text* adalah *Filtering*, yaitu menghapus atau menyimpan pilihan kata. Disini adalah proses dimana proses *filtering* kata-kata, simbol yang tidak diperlukan. *Filtering* disini melibatkan penghapusan atau penyimpanan kata-kata tertentu. Salah satunya adalah penghapusan *stopwords*, yaitu kata-kata penutup yang umumnya tidak memberikan kontribusi signifikan dalam analisis teks. Selain itu, *filtering* juga dapat memuat daftar kata-kata khusus yang disediakan pengguna dalam format file *txt* untuk menghilangkan kata yang tidak diperlukan, dengan satu *stopword* per baris.



Gambar. 3. Design Widget Preprocess Text

Data yang sudah dilakukan *preprocess text* kemudian disimpan dalam bentuk *csv*, dimana file *csv* ini akan digunakan untuk proses analisis sentimen.



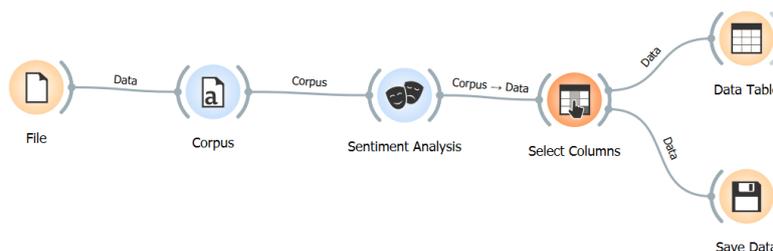
Gambar. 4. Design Widget Simpan Data Twitter

Setelah melakukan *preprocess text* dan menyimpan data yang terkumpul, hasilnya dapat divisualisasikan dalam bentuk *word cloud* menggunakan *Orange Data Mining* seperti yang terlihat pada Gambar 5. Data yang sudah dilakukan *preprocess text* tersisa 1022 *tweet* dari data awal.



Gambar. 5. Word Cloud Identitas Kependudukan Digital

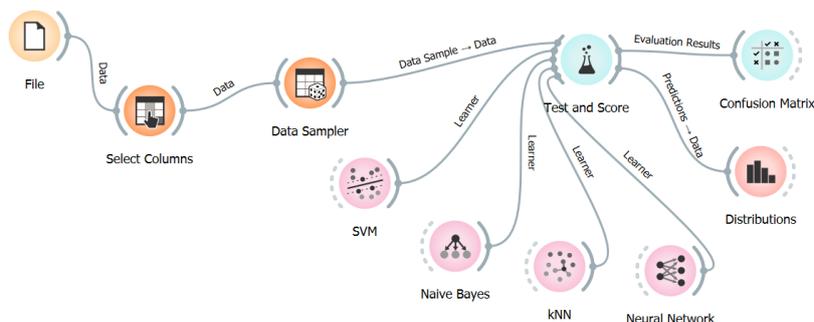
Data *tweet* yang telah disimpan sebelumnya diproses untuk melakukan analisis sentimen, dan kolom yang dipilih untuk tahap selanjutnya ditampilkan menggunakan *Data Table* dalam *widget select column* [15]. Dalam tahapan analisis sentimen ini, penulis menggunakan metode *Multilingual Sentiment* karena *tweet* yang digunakan menggunakan bahasa Indonesia. Pada tahap ini, dilakukan *labeling* untuk analisis positif, negatif, maupun netral. Hasil dari analisis disimpan dalam format *csv* untuk kemudian dilakukan pengecekan *Test and Score* guna membandingkan metode klasifikasi.



Gambar. 6. Design Widget Analisis Sentimen

D. Modelling

Pada langkah ini, sesuai dengan pendekatan algoritma yang dibahas, akan digunakan metode K-NN, *Naive Bayes*, SVM, dan *Neural Network*. Gambar 7 akan menunjukkan tampilan aplikasi *Orange Data Mining* dengan ketiga model tersebut. Pada Gambar 7, model klasifikasi dapat dipasang secara bersamaan dengan metode K-NN, *Naive Bayes*, SVM, dan *Neural Network* yang tersedia dalam aplikasi *Orange Data Mining*. Data yang telah diimpor dari file *csv* setelah proses *crawling* menggunakan API, dan telah melalui tahap pembersihan, akan ditampilkan dalam Data Tabel. Data tersebut akan digunakan sebagai *input* untuk melatih dan menguji model pada Model *Test and Score*. Tujuan dari langkah ini adalah untuk mendapatkan nilai akurasi, presisi, dan *recall* dari model yang telah dilatih dan diuji menggunakan data tersebut.



Gambar. 7. Model Klasifikasi di *Orange Data Mining*

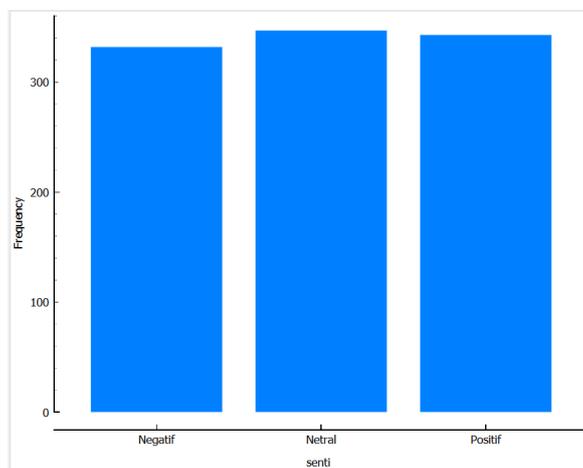
Setelah melakukan pengolahan menggunakan *Orange Data Mining*, diperoleh hasil *Test and Score* dari masing-masing model dengan menggunakan *data training* seperti yang terlihat pada Gambar 8. Berdasarkan pengujian terhadap 1022 data, dilakukan perhitungan presisi (*precision*), *recall*, dan akurasi (*accuracy*) untuk setiap model. Hasil klasifikasi K-NN menunjukkan nilai akurasi tertinggi dibandingkan dengan model lainnya. Dalam perbandingan 4 model, nilai AUC (*Area Under the Curve*) tertinggi diperoleh oleh metode K-NN, yaitu sebesar 1. Sementara itu, metode *Neural Network* dan *Naive Bayes* menghasilkan nilai yang relatif serupa untuk ketiga indikator penilaian. Namun, metode SVM menunjukkan nilai yang paling rendah dalam hal akurasi. AUC digunakan sebagai ukuran kinerja diskriminatif dengan memperkirakan probabilitas *output* dari sampel yang dipilih secara acak dari populasi positif dan negatif. Semakin besar nilai AUC, semakin baik hasil klasifikasi yang digunakan.

| Model | AUC | CA | F1 | Precision | Recall |
|----------------|-------|-------|-------|-----------|--------|
| kNN | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| SVM | 0.818 | 0.448 | 0.394 | 0.468 | 0.448 |
| Neural Network | 0.964 | 0.941 | 0.941 | 0.942 | 0.941 |
| Naive Bayes | 0.963 | 0.923 | 0.921 | 0.936 | 0.923 |

Gambar. 8. Hasil Widget *Test and Score*

Proses *labeling* dan analisis sentimen telah selesai dilakukan, langkah selanjutnya adalah menghubungkan data tersebut ke komponen *Distribution* dalam *Orange Data Mining*. Pada tahap ini, hasil analisis sentimen akan digunakan untuk mengidentifikasi bentuk sentimen dalam data twitter yang telah diolah. Data akan dianalisis dan divisualisasikan dalam bentuk sentimen positif, negatif, atau netral melalui komponen *Distribution*. Ini akan membantu dalam pemahaman visual terhadap distribusi sentimen dari data twitter yang telah diproses.

Berdasarkan hasil yang diperoleh dari analisis data Twitter, pada Gambar 9 terlihat bahwa respon dari masyarakat terhadap topik Identitas Kependudukan Digital cenderung lebih dominan dalam kategori sentimen netral. Distribusi sentimen dalam bentuk visualisasi tidak menunjukkan perbedaan yang signifikan antara sentimen positif, negatif, dan netral. Namun, dari hasil akhir visualisasi tersebut, terlihat bahwa masyarakat yang bersentimen netral lebih mendominasi dalam memberikan tanggapan terhadap topik tersebut. Hal ini menunjukkan bahwa mayoritas tanggapan atau opini yang diungkapkan terhadap topik Identitas Kependudukan Digital tidak memiliki kecenderungan yang jelas ke arah positif atau negatif.



Gambar. 9. Distribution

E. Evaluation

Confusion Matrix adalah metode evaluasi performa untuk masalah klasifikasi dalam *machine learning*, dimana *output* dapat berupa dua kelas atau lebih. Evaluasi hasil untuk setiap model klasifikasi dapat dilihat pada Gambar 10, dimana gambar tersebut menampilkan gabungan *Confusion Matrix* dari model SVM, K-NN, *Naive Bayes*, dan *Neural Network*. Dalam proses evaluasi, penulis melakukan perhitungan manual untuk *precision*, *recall*, dan akurasi dari data yang tersedia. Perbandingan hasil perhitungan manual ini disajikan dalam bentuk tabel.

| SVM | | | | Naive Bayes | | | |
|---------|-----------|--------|---------|-------------|-----------|--------|---------|
| | Predicted | | | | Predicted | | |
| | Negatif | Netral | Positif | | Negatif | Netral | Positif |
| Negatif | 92 | 237 | 3 | Negatif | 254 | 78 | 0 |
| Netral | 21 | 27 | 299 | Netral | 1 | 346 | 0 |
| Positif | 4 | 0 | 339 | Positif | 0 | 0 | 343 |
| Σ | 117 | 264 | 641 | Σ | 255 | 424 | 343 |

| K-NN | | | | Neural Network | | | |
|---------|-----------|--------|---------|----------------|-----------|--------|---------|
| | Predicted | | | | Predicted | | |
| | Negatif | Netral | Positif | | Negatif | Netral | Positif |
| Negatif | 332 | 0 | 0 | Negatif | 314 | 16 | 2 |
| Netral | 0 | 347 | 0 | Netral | 14 | 324 | 9 |
| Positif | 0 | 0 | 343 | Positif | 3 | 16 | 324 |
| Σ | 332 | 347 | 343 | Σ | 331 | 356 | 335 |

Gambar. 10. Confusion Matrix Keempat Metode

Berdasarkan Tabel 1 dibawah, metode K-NN menunjukkan kinerja yang sangat baik dalam klasifikasi data *tweet* terkait Identitas Kependudukan Digital. Akurasinya mencapai hasil yang sangat tinggi, bahkan terbilang sempurna. Namun, penting untuk mempertimbangkan apakah hasil yang sempurna ini juga terjadi pada *dataset* yang lebih besar atau jika ada masalah khusus yang tidak dihadapi dalam penelitian ini. Hasil penelitian ini berbeda dengan penelitian yang dilakukan oleh E. F. Wati dan B. Rudianto (2022), penelitian tersebut mempunyai hasil nilai AUC yang sangat rendah untuk kinerja KNN [16]. Keberhasilan dari penelitian ini dipengaruhi oleh kualitas data uji dan data latih yang digunakan dalam proses simulasi.

TABEL 1
PERBANDINGAN KINERJA

| Metode | Accuracy | Precision | Recall |
|----------------|----------|-----------|--------|
| K-NN | 100% | 100% | 100% |
| SVM | 45% | 45% | 47% |
| Neural Network | 94% | 94% | 94% |
| Naive Bayes | 93% | 92% | 94% |

Metode Naive Bayes memiliki nilai akurasi yang hampir sama dengan *Neural Network* dalam penelitian ini, namun nilai akurasi tertinggi setelah K-NN diperoleh dengan menggunakan metode *Neural Network*. Ini menunjukkan bahwa *Naive Bayes* dapat menjadi alternatif yang baik dalam masalah klasifikasi, namun performa *Naive Bayes* dapat bervariasi tergantung pada karakteristik data yang digunakan. *Neural Network* umumnya dapat memberikan hasil yang baik dalam masalah klasifikasi, terutama ketika memiliki struktur yang tepat dan diatur dengan baik. Penelitian ini selaras dengan penelitian sebelumnya yang dilakukan oleh Ridwan dan tim (2020) bahwa penerapan *Neural Network* memiliki akurasi yang mendekati angka 100% untuk memprediksi kelulusan mahasiswa manajemen di Universitas Bhayangkara Jakarta Raya *University* dengan sampel angkatan 2014 dan 2015 [17].

Sedangkan untuk metode SVM memberikan nilai perhitungan yang paling kecil dan cukup berbeda dengan tiga metode lainnya yang dibandingkan dalam penelitian ini. Meskipun SVM merupakan metode klasifikasi yang kuat, seperti yang dijelaskan pada penelitian oleh D. Derisma dan F. Febrian (2020) dengan nilai akurasi hampir sempurna dalam mendeteksi kanker payudara, hasil ini menunjukkan bahwa dalam konteks penelitian terkait Identitas Kependudukan Digital, SVM tidak seefektif metode lainnya dalam klasifikasi data *tweet* [4]. Hal ini mengindikasikan bahwa metode K-NN memiliki potensi untuk memberikan hasil yang lebih baik dalam konteks penelitian yang berbeda, meskipun metode SVM tetap memiliki keunggulan dalam menangani data dengan adanya *noise*.

F. Deployment

Evaluasi mendalam terhadap hasil model telah selesai dilakukan, langkah selanjutnya adalah mengimplementasikan keseluruhan model yang telah dikembangkan. Dalam implementasi, penting untuk memastikan bahwa model dapat diintegrasikan dengan sistem atau proses yang ada, dan mampu menghasilkan hasil yang relevan dan bermanfaat sesuai dengan tujuan awal. Selama tahap ini, juga dilakukan penyesuaian terhadap model untuk memastikan bahwa hasil yang dihasilkan sesuai dengan target awal dalam tahap CRISP-DM ini.

IV. KESIMPULAN

Dari hasil pengujian untuk membandingkan metode SVM, K-NN, *Naive Bayes*, dan *Neural Network*, serta mengetahui sentimen dan performa terbaik, ditemukan bahwa masyarakat cenderung memiliki sentimen netral terkait Identitas Kependudukan Digital. Penelitian ini memberikan peluang bagi pemerintah untuk mengembangkan inovasi di era digital dan mempermudah akses terhadap data kependudukan. Hasil penelitian menunjukkan bahwa metode K-NN merupakan metode terbaik untuk menganalisa sentimen terkait Identitas Kependudukan Digital dengan nilai akurasi, presisi, dan *recall* mencapai 100%. Metode *Neural Network* dan *Naive Bayes* memiliki perbedaan yang kecil dalam performanya, sedangkan metode SVM memiliki performa yang lebih rendah. Sebagai saran pengembangan, penelitian selanjutnya dapat memperbanyak jumlah data dan mempertimbangkan teknik pengolahan data yang lebih canggih serta eksplorasi metode klasifikasi lainnya untuk meningkatkan kinerja model.

DAFTAR PUSTAKA

- [1] Dukcapiladmin, "Identitas Kependudukan Digital (IKD) / Digital ID," *Disdukcapil Madiun*, 2023. <https://dukcapil.madiunkab.go.id/identitas-kependudukan-digital-digital-id/> (accessed May 05, 2023).
- [2] D. D. PURNAMASARI, "Pengguna Identitas Kependudukan Digital Ditarget Capai 50 Juta Jiwa di 2023," *Kompas*, 2023. <https://www.kompas.id/baca/polhuk/2023/05/03/pengguna-identitas-kependudukan-digital-ditarget-capai-50-juta-jiwa-di-2023> (accessed May 06, 2023).
- [3] S. Indah Nurhafida and F. Sembiring, "Analisis Text Clustering Masyarakat Di Twitter Mengenai Mcdonald'Sxmts Menggunakan Orange Data Mining," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.*, pp. 28–35, 2021.
- [4] D. Derisma and F. Febrian, "Perbandingan Teknik Klasifikasi Neural Network, Support Vector Machine, dan Naive Bayes dalam Mendeteksi Kanker Payudara," *Bina Insa. Ict J.*, vol. 7, no. 1, p. 53, 2020, doi: 10.51211/biict.v7i1.1343.
- [5] Shedriko, "STRING (Satuan Tulisan Riset dan Inovasi Teknologi) PERBANDINGAN ALGORITMA SVM DAN KNN DALAM MENGLASIFIKASI KELULUSAN MAHASISWA PADA SUATU MATA KULIAH," vol. 6, no. 2, pp. 115–122, 2021.
- [6] F. S. Ramadhan and A. P. W. Wibowo, "TEXT CLUSTERING ANALYSIS OF PUBLIC SENTIMENT ON TWITTER ABOUT TOKOPEDIA BTS USING ORANGE DATA MINING," *J. DARMA AGUNG*, vol. JURNAL DAR, pp. 784–792, 2022.
- [7] R. A. raffaidy Wiguna and A. I. Rifai, "Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law Menggunakan Orange Data Mining," *J. Inf. Syst. Informatics*, vol. 3, no. 1, pp. 1–12, 2021, doi: 10.33557/journalisi.v3i1.78.
- [8] D. Abimanyu, E. Budianita, E. P. Cynthia, F. Yanto, and Y. Yusra, "Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 3, pp. 423–431, 2022, doi: 10.32672/jnkti.v5i3.4382.
- [9] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART

- untuk Prediksi Curah Hujan Berpotensi Banjir,” *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [10] A. Liken Anggoro, L. V. P. Ken, and M. G. Setiawan, “Analisis Media Text Clustering pada Twitter Akan Kasus Selebriti Menggunakan Orange Data Mining,” *Remik*, vol. 7, no. 1, pp. 189–195, 2023, doi: 10.33395/remik.v7i1.12001.
- [11] H. Hozairi, A. Anwari, and S. Alim, “Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes,” *Netw. Eng. Res. Oper.*, vol. 6, no. 2, p. 133, 2021, doi: 10.21107/nero.v6i2.237.
- [12] N. Anggraini and H. Suroyo, “Comparison of Sentiment Analysis against Digital Payment ‘T-cash and Go-pay’ in Social Media Using Orange Data Mining Perbandingan Analisis Sentimen Terhadap Digital Payment ‘ T-cash dan Go-pay’ Di Sosial Media Menggunakan Orange Data Mining,” *J. Inf. Syst. Informatics*, vol. 1, no. 1, pp. 152–163, 2019, [Online]. Available: <http://journal-isi.org/index.php/isi>
- [13] A. Patunru, A. Surianta, and A. Tenggara, “Menarik Penanaman Modal Asing (PMA),” no. 4, pp. 1–9, 2020.
- [14] Ismail, H. N. Rahmah, and R. Sulistyowati, “Penggunaan Software Orange Data Mining Pada Implementasi Text Mining Dalam Analisis Sentimen Netizen Di Twitter Terhadap Kelangkaan Minyak Goreng Implementation of Text Mining in Netizen Sentiment Analysis on Twitter on Cooking Oil Scarcity Using Orange D,” *Sigma-Mu*, vol. 14, no. 2, pp. 1–11, 2022.
- [15] N. A. Sinaga, B. H. Hayadi, and Z. Situmorang, “Perbandingan Akurasi Algoritma Naive Bayes, K-Nn Dan Svm Dalam Memprediksi Penerimaan Pegawai,” *J. Tek. Inf. dan Komput.*, vol. 5, no. 1, p. 27, 2022, doi: 10.37600/tekinkom.v5i1.446.
- [16] E. F. Wati and B. Rudianto, “Penerapan Algoritma KNN, Naive Bayes Dan C4.5 Dalam Memprediksi Kelulusan Mahasiswa” *Jurnal Format*, vol. 11, no. 2, pp. 168–175, 2022.
- [17] R. Ridwan, H. Lubis, and P. Kustanto, “Implementasi Algoritma Neural Network dalam Memprediksi Tingkat Kelulusan Mahasiswa,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 286, 2020, doi: 10.30865/mib.v4i2.2035.