

ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP PELAYANAN DINAS KEPENDUDUKAN DAN PENCATATAN SIPIL MENGGUNAKAN ALGORITMA NAÏVE BAYES

Emil Salim¹, Achmad Solichin^{2*}

^{1,2}Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia
Email: ¹1911510079@student.budiluhur.ac.id, ^{2*}achmad.solichin@budiluhur.ac.id
(* : coresponding author)

Abstrak-Dinas Kependudukan dan Pencatatan Sipil (Disdukcapil) memiliki tugas utama dalam memberikan layanan kependudukan dan pencatatan sipil bagi masyarakat. Saat ini, di kalangan masyarakat masih sering memiliki stigma kurang baik terhadap proses pengurusan dokumen kependudukan yang terkesan lama, rumit dan sering terjadi pungutan liar oleh pihak-pihak tertentu. Pada penelitian ini dilakukan analisis sentimen masyarakat terhadap layanan Disdukcapil pada media sosial Twitter. Analisis sentimen diperoleh menggunakan pendekatan machine learning disertai dengan ekstraksi fitur menggunakan TF-IDF dan algoritma Naïve Bayes. Dataset yang digunakan pada penelitian ini diperoleh dari media sosial Twitter berupa kicauan (tweet) berbahasa Indonesia. Dataset diperoleh menggunakan kata kunci 'dukcapil', 'kependudukan', 'disdukcapil', 'ngurus ktp', 'pembuatan ktp', 'pengalaman ktp', 'pelayanan disdukcapil', 'ngurus kk', 'pelayanan ktp', 'ktp rusak', 'dokumen kependudukan', 'ngurus disdukcapil', 'bikin kartu keluarga', 'pelayanan dukcapil', 'bikin akta lahir', 'buat akta lahir', dan 'disdukcapil ramah'. Hasil analisis berdasarkan 647 tweet menunjukkan bahwa sentimen positif sebesar 70,20% dan sentimen negatif sebanyak 29,80%, dengan nilai akurasi 65,39% presisi 97,65%, recall 66,87%, dan f1-score sebesar 39,69%. Hasil penelitian ini bermanfaat bagi Disdukcapil untuk meningkatkan kualitas layanannya sehingga sentimen masyarakat semakin baik.

Kata Kunci: analisis sentimen, media sosial twitter, layanan kependudukan, disdukcapil, naïve bayes

Abstract-The Department of Population and Civil Registration (Disdukcapil) has the main task of providing population and civil registration services for the community. Currently, the community still often has a bad stigma about the process of managing population documents which seems long, complicated and often illegal levies occur by certain parties. In this study, an analysis of public sentiment towards the Disdukcapil service was carried out on Twitter social media. Sentiment analysis was obtained using a machine learning approach accompanied by feature extraction using TF-IDF and the Naïve Bayes algorithm. The dataset used in this study was obtained from Twitter social media in the form of Indonesian-language tweets. The dataset was obtained using the keywords 'dukcapil', 'population', 'disdukcapil', 'ngurus ktp', 'ktp making', 'ktp experience', 'disdukcapil service', 'take care of family card', 'ktp service', 'broken ID card', 'residency document', 'manage disdukcapil', 'make a family card', 'dukcapil service', 'make a birth certificate', 'make a certificate born', and 'friendly disdukcapil'. The results of the analysis based on 647 Tweets show that positive sentiment is 70.20% and negative sentiment is 29.80%, with an accuracy value of 65.39%, precision of 97.65%, recall of 66.87, and f1-score of 39.69%. The results of this study are useful for Disdukcapil to improve the quality of its services so that public sentiment is getting better.

Keywords: sentiment analysis, twitter sosial media, civil registration service, disdukcapil, naïve bayes

1. PENDAHULUAN

Dinas Kependudukan dan Pencatatan Sipil merupakan unsur pelaksana Pemerintah Daerah di bidang Kependudukan dan Pencatatan Sipil yang dipimpin oleh Kepala Dinas dan berkedudukan di bawah dan bertanggung jawab kepada Bupati melalui Sekretaris Daerah. Dinas Kependudukan dan Pencatatan Sipil mempunyai tugas melaksanakan urusan rumah tangga Pemerintah Daerah dan tugas pembantuan di bidang Kependudukan dan Pencatatan Sipil. Dinas Kependudukan dan Pencatatan Sipil memiliki fungsi merumuskan kebijakan bidang pendaftaran penduduk, pencatatan sipil, dan pengelolaan sistem informasi administrasi kependudukan, melaksanakan evaluasi dan pelaporan bidang pendaftaran penduduk, pencatatan sipil, dan pengelolaan sistem informasi administrasi kependudukan, melaksanakan administrasi dinas sesuai dengan lingkup tugasnya dan melaksanakan fungsi lain yang diberikan oleh Bupati sesuai dengan tugas dan fungsinya [1]. Pelayanan kepada masyarakat tentunya menjadi kewajiban Dinas Kependudukan dan Pencatatan Sipil yang mendapatkan berbagai tanggapan juga pendapat dari masyarakat. Pendapat yang disampaikan ada yang baik dan membangun, tetapi ada juga yang kurang merasa puas dengan pelayanan yang diberikan atau bertolak belakang dengan harapan.

Media sosial Twitter menjadi salah satu tempat masyarakat dapat bebas menyampaikan pendapat. Banyak sekali metode analisis yang bisa digunakan untuk menganalisis pendapat masyarakat berdasarkan informasi yang ada pada media sosial semacam Twitter. Salah satu yang ada adalah analisis sentimen. Metode analisis sentimen merupakan salah satu metode untuk menganalisis data yang didapatkan dari internet sehingga bisa diketahui atau dikutip sebuah informasi polaritas dari data tersebut. Dengan menggunakan analisis sentimen polaritas dari opini

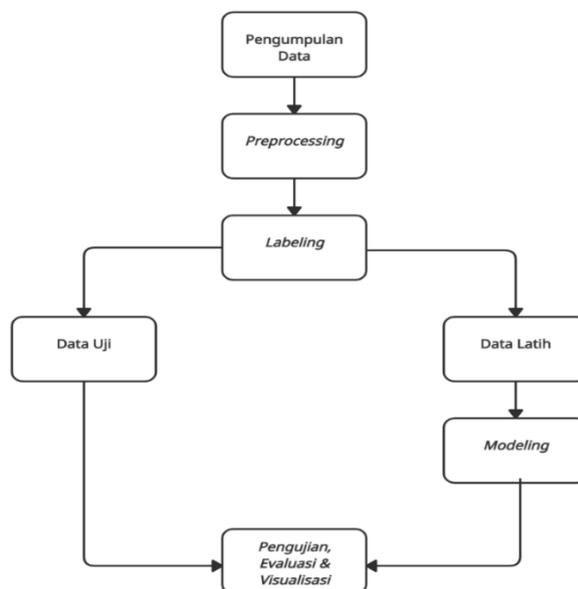
yang ada dapat dikumpulkan, sehingga dapat digunakan untuk memprediksi masyarakat puas atau tidak puas nya pada pelayanan Dinas Kependudukan dan Pencatatan Sipil. Beberapa penelitian tentang analisis sentimen ini pada konten media sudah dilakukan sebelumnya. Pada penelitian ini, dilakukan prediksi analisis sentimen kepuasan pelayanan masyarakat pada Dinas Kependudukan dan Pencatatan Sipil dengan menggunakan algoritme Naïve Bayes.

Pada penelitian oleh Setian dan Seprina [2] yang melakukan analisis sentimen masyarakat terhadap layanan e-commerce Lazada disimpulkan bahwa algoritme Naïve Bayes berhasil diterapkan untuk mengklasifikasikan data tweet dengan akurasi sebesar 98,29%. Penelitian lain yang berjudul Analisis Sentimen Physical Distancing pada Twitter Menggunakan Text Mining dengan Algoritme Naïve Bayes Clasifier [3], menyatakan bahwa Algoritme Naïve Bayes pada penelitiannya mampu meraih akurasi sebesar 50,26%. Pada penelitian lainnya juga yaitu Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Pada Angket Mahasiswa [4], didapat nilai akurasi sebesar 80% dan tingkat presisi 75% recall sebesar 75%. Penelitian lainnya melakukan analisis sentimen masyarakat terhadap pembelajaran daring menggunakan metode KNN dengan nilai akurasi sebesar 71% [5].

Pada penelitian ini dilakukan analisis sentimen masyarakat terhadap layanan Dinas Kependudukan dan Pencatatan Sipil menggunakan pendekatan metode Naïve Bayes. Dataset yang digunakan berupa teks kicauan (*tweet*) dengan kata kunci, ‘dukcapil’, ‘kependudukan’, ‘disdukcapil’, ‘ngurus ktp’, ‘pembuatan ktp’, ‘pengalaman ktp’, ‘pelayanan disdukcapil’, ‘ngurus kk’, ‘pelayanan ktp’, ‘ktp rusak’, ‘dokumen kependudukan’, ‘ngurus disdukcapil’, ‘bikin kartu keluarga’, ‘pelayanan dukcapil’, ‘bikin akta lahir’, ‘buat akta lahir’, dan ‘disdukcapil ramah’, sehingga dapat mengetahui seberapa besar sentimen positif atau negatif masyarakat terhadap pelayanan Dinas Kependudukan dan Pencatatan Sipil. Dari penelitian ini, diharapkan dapat menjadi salah satu pengetahuan yang dapat dijadikan sebagai acuan perubahan atau peningkatan pelayanan Dinas Kependudukan dan Pencatatan Sipil.

2. METODE PENELITIAN

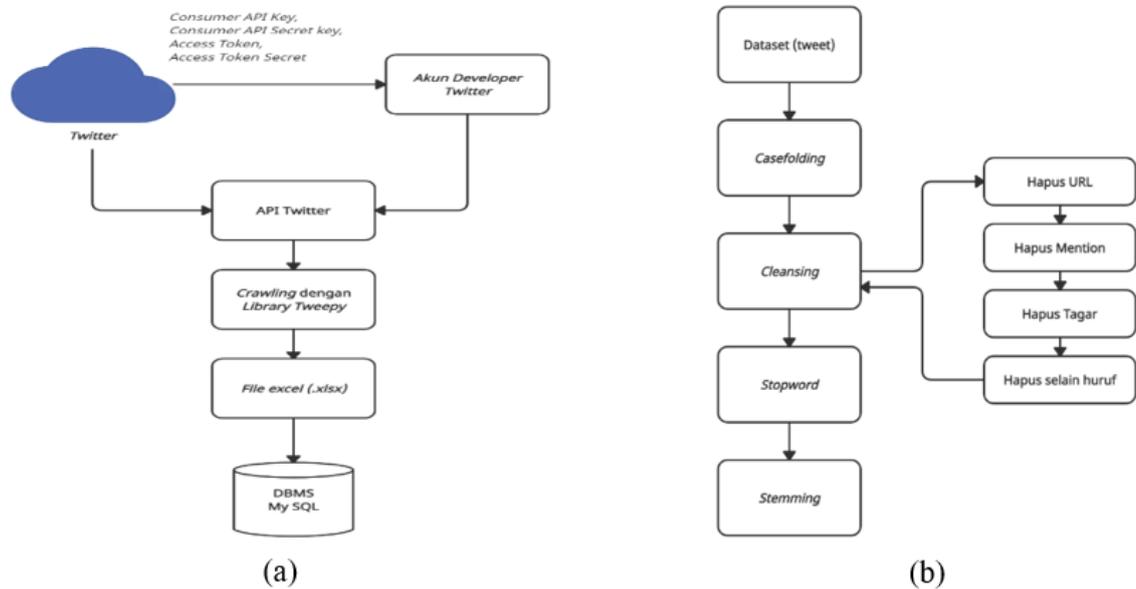
Pengembangan sistem analisis sentimen yang dilakukan pada penelitian ini melibatkan beberapa tahapan. Gambar 1 menyajikan tahapan penelitian yang menjadi acuan dalam pelaksanaan penelitian ini. Penelitian diawali dengan pengumpulan data yang dilanjutkan dengan tahap pengolahan awal (*preprocessing*), pelabelan, hingga evaluasi.



Gambar 1. Tahapan Penelitian

Pengumpulan data dilakukan melalui proses *crawling* untuk mendapatkan *dataset* berupa kicauan (*tweet*) yang selanjutnya diolah menggunakan perangkat lunak *Microsoft Excel* dan dimasukkan ke dalam pangkalan data. Gambar 2 bagian (a) menyajikan tahap pengumpulan data kicauan (*tweet*) yang memanfaatkan API yang disediakan oleh Twitter. *Twitter API key* diperoleh melalui akun pengembang Twitter (<https://developer.twitter.com/>). Beberapa komponen Twitter API Key yang diperoleh meliputi: kunci API Klien, kunci rahasia API Klien, token

akses, dan rahasia token akses. Proses selanjutnya adalah menambang data yang berasal dari media sosial Twitter menggunakan *library* Tweepy di Python. Data *tweet* yang dikumpulkan akan disimpan dalam file Excel (.xlsx) dan kemudian disimpan ke sistem pangkalan data MySQL.



Gambar 2. Tahap penelitian: (a) pengumpulan data, dan (b) preprocessing data

Selanjutnya, pada proses *preprocessing* dilakukan penyaringan, pembuangan dan perbaikan kata. Hasil dari proses *preprocessing* menghasilkan teks yang lebih terstruktur (*clean text*) yang digunakan pada tahap selanjutnya. Tahap *preprocessing* terdiri dari beberapa proses antara lain *case folding*, *cleansing*, menghapus *stop word*, dan proses *stemming*. Tahap *preprocessing* yang dilakukan pada penelitian ini disajikan pada Gambar 2 bagian (b).

Proses *case folding* merupakan proses mengubah teks menjadi huruf kecil (*lowercase*) atau huruf besar (*uppercase*) sehingga dihasilkan teks yang seragam. Dalam penelitian ini, proses *case folding* mengubah teks cuitan (*tweet*) menjadi huruf kecil semua. Sebagai contoh, jika terdapat *tweet* yang mengandung kata “Dukcapil” atau “DUKCAPIL” maka keduanya akan diubah menjadi “dukcapil”. Tahap selanjutnya adalah proses *cleansing*. Pada proses *cleansing*, teks dibersihkan dari karakter atau unsur lain yang tidak diperlukan atau tidak bermakna. Pada penelitian ini, proses *cleansing* terdiri dari beberapa tahap yaitu menghapus alamat URL, menghapus tanda mention (@), menghapus tagar (#) dan menghapus karakter selain huruf a-z.

Selanjutnya, proses penghapusan kata dalam teks yang perlu diabaikan karena merupakan kata sambung atau kurang memiliki makna. Pada penelitian ini, digunakan daftar *stopword* yang dikembangkan oleh [6], [7]. Beberapa contoh yang termasuk dalam daftar *stopword* antara lain kata “untuk”, “dan”, “bagi”. Kata-kata yang masuk dalam daftar *stopword* tersebut dihilangkan atau dihapus. Tahap *preprocessing* selanjutnya adalah tahap *stemming*. Proses *stemming* merupakan proses mengubah kata berimbuhan menjadi kata dasar dengan melibatkan pustaka Sastrawi yang dikembangkan dari penelitian Nazief-Adriani [6], misalnya kata ‘membantu’, menjadi ‘bantu’, ‘pembelajaran’ menjadi ‘ajar’. Proses pengubahan tersebut dilakukan dengan melibatkan Pustaka Sastrawi, menggunakan paket *StemmerFactory* di Python.

Tabel 1. Contoh Penerapan Proses Preprocessing Data

Tahap	Teks Awal	Teks Setelah Proses
<i>Case folding</i>	Birokrasi pelayanan saat ini untuk masyarakat, sudah nyaman loh https://twitter.com #jakarta	birokrasi pelayanan saat ini untuk masyarakat, sudah nyaman loh https://twitter.com #jakarta
Cleansing: menghapus URL	birokrasi pelayanan saat ini untuk masyarakat, sudah nyaman loh https://twitter.com #jakarta	birokrasi pelayanan saat ini untuk masyarakat, sudah nyaman loh #jakarta
Cleansing: menghapus mention @	birokrasi pelayanan saat ini untuk masyarakat, sudah nyaman loh @ #jakarta	birokrasi pelayanan saat ini untuk masyarakat, sudah nyaman loh #jakarta

Cleansing: menghapus tagar #	birokrasi pelayanan saat ini untuk masyarakat, sudah nyaman loh #jakarta	birokrasi pelayanan saat ini untuk masyarakat, sudah nyaman loh jakarta
Cleansing: menghapus karakter selain a-z	birokrasi pelayanan saat ini untuk masyarakat, sudah nyaman loh jakarta	birokrasi pelayanan saat ini untuk masyarakat sudah nyaman loh jakarta
Menghilangkan stopword	birokrasi pelayanan saat ini untuk masyarakat sudah nyaman loh jakarta	birokrasi pelayanan masyarakat nyaman jakarta
Stemming	birokrasi pelayanan masyarakat nyaman jakarta	birokrasi layan masyarakat nyaman jakarta

Clean text yang diperoleh dari proses *preprocessing* diproses dalam tahap pelabelan untuk menentukan kelas atau label. Dalam penelitian ini, digunakan 2 (dua) kelas sentimen, yaitu *positif* atau *negatif*. Pelabelan dilakukan secara manual oleh pakar yang menilai sentimen positif atau negatif dari setiap *tweet*. Setelah pelabelan dilakukan, selanjutnya dataset dibagi menjadi 2 (dua) bagian, yaitu data latih (*training*) dan data uji (*testing*). Komposisi yang digunakan adalah 80% data latih dan 20% data uji dengan pemilihan data secara acak proporsional. Hal tersebut dilakukan untuk memastikan keberimbangan antara data label positif dan negatif.

Setelah dilakukan tahap *preprocessing*, selanjutnya dilakukan pembobotan menggunakan algoritma TF-IDF. Term Frequency-Inverse Document Frequency atau TF-IDF merupakan metode yang berguna untuk menghitung bobot setiap kata yang umum digunakan. Metode ini banyak digunakan karena efisien, mudah dan memiliki hasil yang akurat seperti pada [8]–[10]. Metode ini menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. Secara sederhana, metode TF-IDF digunakan untuk mengetahui berapa sering suatu kata muncul di dalam dokumen.

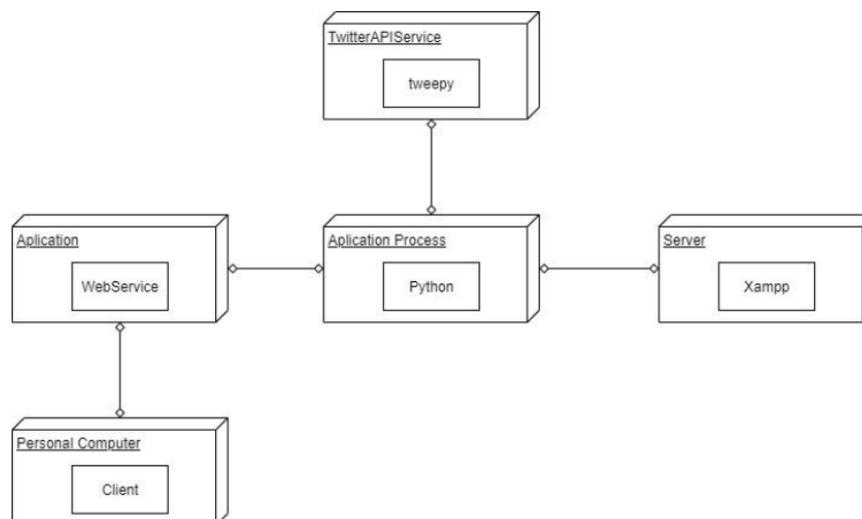
Selanjutnya, metode klasifikasi yang digunakan pada penelitian ini adalah Multinomial Naïve Bayes yang terbukti handal dan efektif untuk klasifikasi data teks, termasuk analisis sentimen [11], [12]. Naïve Bayes merupakan salah satu metode klasifikasi yang didasarkan pada Teorema Bayes. Metode ini pada dasarnya mengklasifikasi data berdasarkan probabilitas kemunculan atribut-atribut terhadap kelas yang diprediksikan. Pada penelitian ini, kelas sentimen terdiri dari 2 (dua) yaitu POSITIF dan NEGATIF.

3. HASIL DAN PEMBAHASAN

3.1. Prototipe Sistem Analisis Sentimen

Pada penelitian ini, dikembangkan prototipe sistem analisis sentimen masyarakat terhadap layanan Disdukcapil dengan mengimplementasikan metode klasifikasi Naïve Bayes. Selain untuk mempermudah analisis sentimen bagi pengguna, pengembangan prototipe bertujuan untuk menguji apakah metode dan algoritma yang diusulkan dapat melakukan klasifikasi sentimen dengan baik. Pada penelitian ini, prototipe dikembangkan dengan bahasa pemrograman Python dan beberapa pustaka pendukungnya. Gambar 3 menyajikan diagram *deployment* dari prototipe sistem analisis sentimen yang dikembangkan pada penelitian ini.

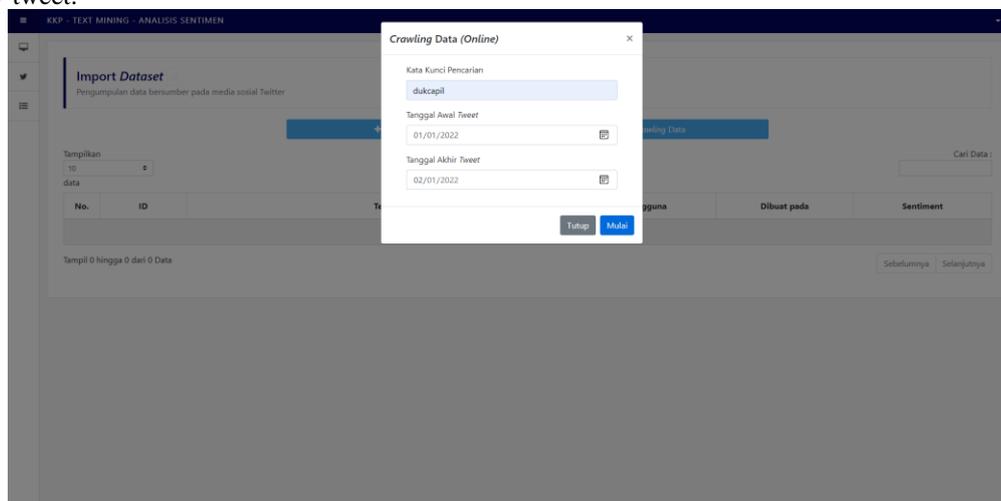
prototipe sistem analisis sentimen yang dikembangkan pada penelitian ini.n pada penelitian ini.



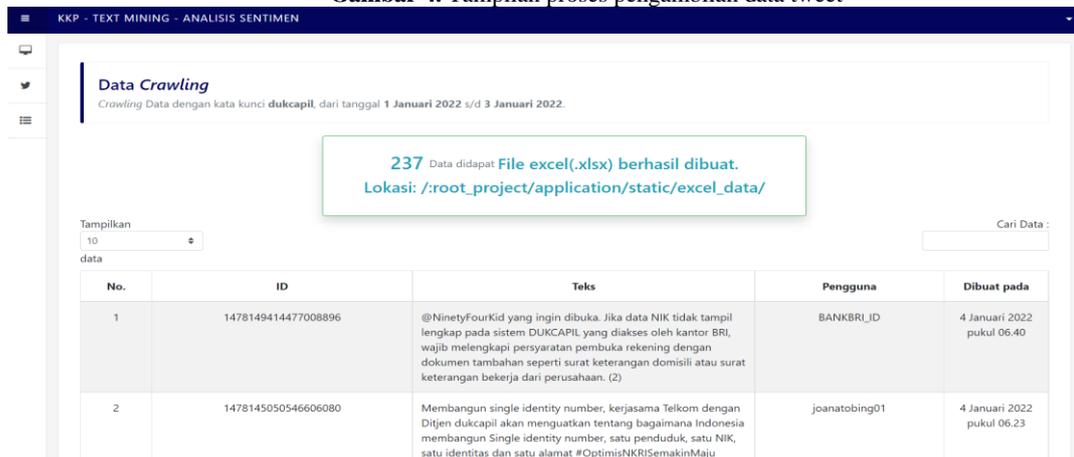
Gambar 3. Deployment Diagram Prototipe Sistem

Untuk menjalankan aplikasi, pengguna melakukan login terlebih dahulu menggunakan username dan password yang telah terdaftar. Setelah login, pengguna dapat melakukan pengambilan data (*crawling*) tweet secara daring menggunakan kata kunci yang telah ditentukan. Pengguna juga dapat mengatur rentang tanggal pengambilan data tweet. Gambar 4 menyajikan tampilan proses pengambilan data tweet yang dapat dilakukan oleh pengguna. Setelah proses pengambilan data berhasil dilakukan, maka akan disajikan notifikasi jumlah data tweet sebagaimana disajikan pada Gambar 5. Hasil proses pengambilan data tersimpan dalam file Microsoft Excel yang selanjutnya dapat dilakukan proses preprocessing. Adapun data tweet yang ditarik terdiri dari 4 atribut yaitu ID, isi tweet, username yang mengirim tweet dan waktu pengiriman tweet.

Sebagaimana dijelaskan pada bagian metode penelitian, dataset yang terkumpul dilakukan tahap persiapan dan pembersihan data. Pada penelitian ini, digunakan 17 kata kunci yaitu: ‘dukcapil’, ‘kependudukan’, ‘disdukcapil’, ‘ngurus ktp’, ‘pembuatan ktp’, ‘pengalaman ktp’, ‘pelayanan disdukcapil’, ‘ngurus kk’, ‘pelayanan ktp’, ‘ktp rusak’, ‘dokumen kependudukan’, ‘ngurus disdukcapil’, ‘bikin kartu keluarga’, ‘pelayanan dukcapil’, ‘bikin akta lahir’, ‘buat akta lahir’, dan ‘disdukcapil ramah’. Setelah dilakukan pembersihan data, diperoleh dataset dengan total data sebanyak 647 tweet.



Gambar 4. Tampilan proses pengambilan data tweet



Gambar 5. Hasil proses pengambilan data tweet tersimpan di file Excel

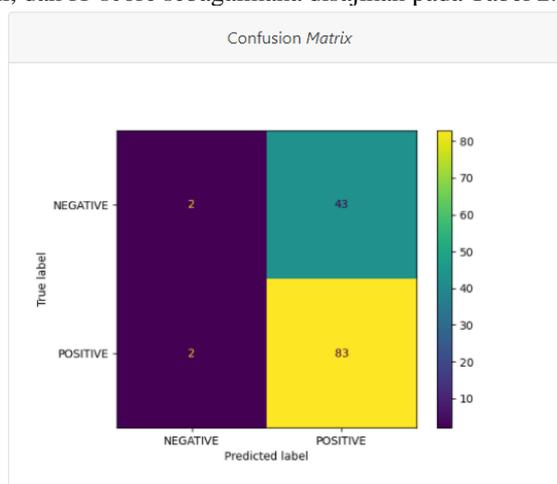
3.2. Pengujian Algoritma

Pada penelitian ini dilakukan pengujian performa algoritma Naïve Bayes dalam mengklasifikasikan sentimen dari tweet yang diuji. Adapun jumlah dataset tweet yang digunakan adalah 647 data yang dibagi menjadi dua bagian: sebanyak 130 (20%) untuk data uji dan 517 (80%) untuk data latih. Pembagian data dilakukan secara acak dengan tetap memperhatikan proporsi jumlah data di tiap kelas (kelas sentimen POSITIF dan NEGATIF).

Sementara itu, untuk ukuran performa algoritma klasifikasi digunakan ukuran akurasi, presisi dan recall yang diperoleh dengan memetakan hasil pengujian menggunakan tabel confusion matrix. Pada tabel confusion matrix

terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah True Positive (TP), True Negative (TN), False Positive (FP) dan False Negative (FN). Nilai True Negative (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan False Positive (FP) merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, True Positive (TP) merupakan data positif yang terdeteksi benar. False Negative (FN) merupakan kebalikan dari True Positive, sehingga data positif, namun terdeteksi sebagai data negatif.

Gambar 6 menyajikan hasil pengujian performa algoritma klasifikasi menggunakan data uji yang berjumlah 130 data. Berdasarkan pemetaan hasil pengujian algoritma klasifikasi menggunakan confusion matrix seperti pada Gambar 6, dapat diperoleh nilai TP = 83, TN = 2, FP = 2, dan FN = 43. Berdasarkan nilai tersebut, selanjutnya dapat diperoleh nilai akurasi, presisi, recall, dan f1-score sebagaimana disajikan pada Tabel 2.

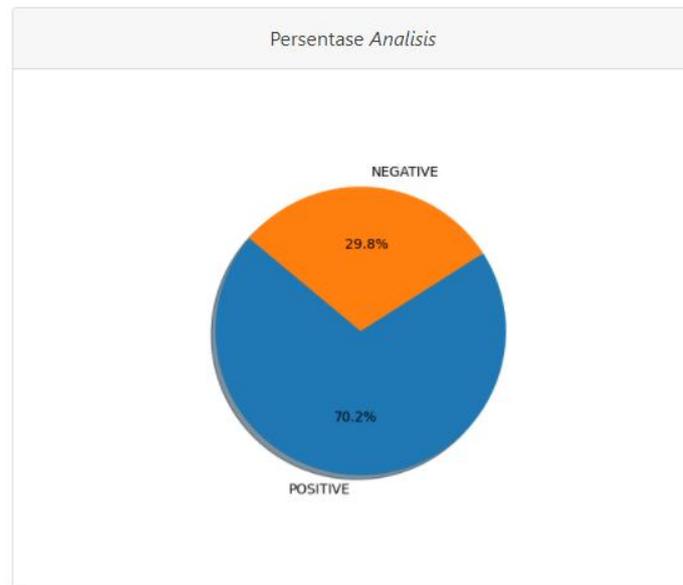


Gambar 6. Hasil Pengujian Performa Algoritma Klasifikasi Menggunakan Confusion Matrix

Tabel 2. Hasil Perhitungan Nilai Akurasi, Presisi, Recall dan F1-Score

Ukuran	Persamaan	Hasil Perhitungan
Akurasi	$Akurasi = \frac{TP + TN}{(TP + TN + FP + FN)} * 100\%$	65,39 %
Presisi	$Presisi = \frac{TP}{(TP + FP)} * 100\%$	97,65%
Recall	$Recall = \frac{TP}{(TP + FN)} * 100\%$	66,87%
F1-Score	$F1 - score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} * 100\%$	39,69%

Berdasarkan hasil perhitungan performa algoritma diperoleh nilai akurasi sebesar 65,39%, presisi 97,65%, recall 66,87%, dan f1-score sebesar 39,69%. Secara umum nilai tersebut masih dalam rentang yang cukup baik. Oleh karena itu, algoritma Naïve Bayes terbukti dapat digunakan untuk melakukan klasifikasi sentimen masyarakat terhadap layanan Dinas Kependudukan dan Catatan Sipil (Disdukcapil). Berdasarkan hasil tersebut, peningkatan performa algoritma masih perlu ditingkatkan, terutama dari sisi akurasi, recall dan f1-score. Peningkatan dapat dilakukan dengan dua langkah, yaitu memperbaiki algoritma klasifikasi yang digunakan, dan meningkatkan kualitas dan kuantitas dataset dengan melibatkan pakar bahasa dan pakar sosial media.



Gambar 7. Hasil Analisis Sentimen Masyarakat Terhadap Layanan Disdukcapil

Selanjutnya, Gambar 7 menyajikan hasil analisis sentimen masyarakat terhadap layanan kependudukan yang dilakukan oleh Disdukcapil. Berdasarkan hasil analisis terhadap data kicauan (tweet) dari masyarakat di tahun 2021, ternyata 70,2% masyarakat memiliki sentimen POSITIF terhadap layanan Disdukcapil, sementara sisanya sebesar 20,8% masyarakat memiliki sentimen NEGATIF. Dari hasil penilaian tersebut, tentunya dapat menjadi masukan bagi pimpinan dan pihak Disdukcapil untuk senantiasa meningkatkan kualitas layanan kependudukan dan catatan sipil bagi masyarakat.

4. KESIMPULAN

Berdasarkan hasil pengujian dan evaluasi dari aplikasi yang dibuat menggunakan Dataset dan Algoritme yang telah diusulkan, maka dapat disimpulkan bahwa berdasarkan 647 Tweet, arah pandangan (sentimen) masyarakat Indonesia terhadap Dinas Kependudukan dan Pencatatan Sipil cenderung kearah sentimen Positif sebesar 70,2 %. Penggunaan ekstraksi fitur TF-IDF dari scikit-learn dan Naïve Bayes dalam melakukan analisis sentimen dapat berjalan dengan baik dengan perolehan nilai akurasi 65,39%. Aplikasi ini diharapkan bisa membantu user untuk mengidentifikasi sebuah atau sekumpulan tweet untuk dapat menyimpulkan sentimen nya dengan menggunakan metode Naïve Bayes.

Adapun saran yang dapat diberikan sebagai pengembangan lebih lanjut untuk aplikasi ini agar dapat berjalan lebih baik lagi antara lain perlu menambahkan kata kunci pencarian *Tweet* sehingga dapat menghasilkan pandangan yang lebih beragam. Selain itu proses pelabelan dapat dilakukan dengan melibatkan pakar bahasa dan pakar sosial media. Pada penelitian ini hanya diujicoba menggunakan satu komposisi data uji dan data latih. Di penelitian selanjutnya dapat dilakukan pembagian yang lebih baik sehingga diharapkan dapat menghasilkan performa metode yang lebih baik lagi.

DAFTAR PUSTAKA

- [1] B. K. Disdukcapil, "Profil Disdukcapil Bantul," 2021.
- [2] D. Setian and I. Seprina, "Analisis Sentimen Masyarakat Terhadap Data Tweet Lazada Menggunakan Text Mining Dan Algoritma Naive Bayes," *Bina Darma Conf. Comput. Sci.*, pp. 998–1004, 2019.
- [3] Hardi, N. Alkahfi, Y. Handayani, P. Gata, W. Firdaus, and M. Rifqi, "Analisis Sentimen Physical Distancing pada Twitter Menggunakan Text Mining dengan Algoritma Naive Bayes Classifier," *Sistemasi*, vol. 10, no. 1, p. 131, 2021.
- [4] Sunardi, A. Fadlil, and Suprianto, "Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Pada Angket Mahasiswa," *Saintekbu*, vol. 10, no. 2, pp. 1–9, 2018.
- [5] M. Priandi and Painem, "Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor," in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 2021, no. September, pp. 311–319.
- [6] M. Adriani, J. Asian, B. Nazief, S. M. M. Tahaghoghi, and H. E. Williams, "Stemming Indonesian: A Confix-Stripping Approach," *ACM Trans. Asian Lang. Inf. Process.*, vol. 6, no. 4, pp. 1–33, 2007.
- [7] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," Universiteit van Amsterdam, 2003.



- [8] N. K. Widyasanti, I. K. G. D. Putra, and N. K. D. Rusjayanthi, "Seleksi Fitur Bobot Kata dengan Metode TFIDF untuk Ringkasan Bahasa Indonesia," *Merpati*, vol. 6, no. 2, pp. 119–126, 2018.
- [9] N. Andriani and A. Wibowo, "Implementasi Text Mining Klasifikasi Topik Tugas Akhir Mahasiswa Teknik Informatika Menggunakan Pembobotan TF-IDF dan Metode Cosine Similarity Berbasis Web," in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 2021, pp. 130–137.
- [10] W. Desena and A. Solichin, "Pencarian Abstrak Tugas Akhir Mahasiswa Berdasarkan Tingkat Kemiripan Menggunakan Algoritma WInnowing dan Jaccard Similarity pada Universitas Budi Luhur," *J. Inform.*, vol. 17, no. 2, pp. 112–122, 2021.
- [11] S. Juanita, "Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes," *J. Media Inform. Budidarma*, vol. 4, no. 3, pp. 552–558, 2020.
- [12] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 4, no. 2, pp. 113–118, 2018.