

ANALISIS SENTIMEN PADA ULASAN PELAYANAN SUKU DINAS KEPENDUDUKAN DAN PENCATATAN SIPIL KOTA ADMINISTRASI JAKARTA BARAT MENGGUNAKAN ALGORITME *K-NEAREST NEIGHBOR*

Emil Salim^{1*}, Mohammad Syafrullah²

^{1,2} Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur.
Email: ¹kemsalim@outlook.com, ² mohammad.syafrullah@budiluhur.ac.id

(Naskah masuk: 22 Maret 2023, diterima untuk diterbitkan: 30 Mei 2023)

Abstrak

Administrasi kependudukan memiliki peranan yang penting di Indonesia karena terlibat dalam berbagai kegiatan. Di tingkat Pemerintah Daerah, Dinas Kependudukan dan Pencatatan Sipil bertanggung jawab atas hal tersebut. Di Jakarta Barat, Suku Dinas Kependudukan dan Pencatatan Sipil merupakan badan pemerintah yang bertugas secara teknis dalam administrasi kependudukan. Berdasarkan data tahun 2020, jumlah penduduk di Jakarta Barat mencapai 2.434.511 orang. Suku Dinas tersebut bertanggung jawab dalam memberikan pelayanan kepada masyarakat, yang melibatkan berbagai respon dan pendapat dari masyarakat. Beberapa pendapat bersifat positif dan membangun, sementara ada yang merasa tidak puas dengan pelayanan yang diberikan atau tidak sesuai harapan. Penelitian ini bertujuan untuk menganalisis pandangan masyarakat terhadap pelayanan administrasi Suku Dinas Kependudukan dan Pencatatan Sipil di Kota Administrasi Jakarta Barat menggunakan pendekatan machine learning, dengan fitur ekstraksi *CountVectorizer* dan algoritma *K-Nearest Neighbor*. Dataset yang digunakan berasal dari formulir daring yang dapat diakses oleh masyarakat. Hasil analisis dari 386 ulasan menunjukkan bahwa terdapat 311 (80,57%) ulasan positif dan 75 (19,43%) ulasan negatif dengan pengujian terbaik dilakukan dengan menggunakan nilai $K=3$, dengan tingkat akurasi sebesar 83%, presisi 82%, dan *recall* 100%.

Kata kunci: *analisis sentimen, dukcapil, ktp, pelayanan, adminduk, administrasi kependudukan, countvectorizer, k-nearest neighbor.*

SENTIMENT ANALYSIS ON THE REVIEWS OF CIVIL REGISTRATION AND POPULATION SERVICE SUB-DISTRICT OF WEST JAKARTA USING *K-NEAREST NEIGHBOR* ALGORITHM

Abstract

Population administration plays an important role in Indonesia as it is involved in various activities. At the regional government level, the Population and Civil Registration Office is responsible for these matters. In West Jakarta, the Sub-district Population and Civil Registration Office is a government body that handles the technical aspects of population administration. Based on data from 2020, the population in West Jakarta reached 2,434,511 people. This office is responsible for providing services to the community, which involves various responses and opinions from the public. Some opinions are positive and constructive, while others may be dissatisfied with the services provided or have unmet expectations. This research aims to analyze public perspectives on the administration services of the Sub-district Population and Civil Registration Office in West Jakarta using machine learning approach, specifically the *CountVectorizer* feature extraction and *K-Nearest Neighbor* algorithm. The dataset used is sourced from an online form accessible by society. The analysis results from 386 reviews show that there are 311 (80.57%) positive reviews and 75 (19.43%) negative reviews, with the best testing performed using a K value of 3, achieving an accuracy rate of 83%, precision of 82%, and recall of 100%.

Keywords: *sentiment analysis, civil registry and identification agency (dukcapil), id card (ktp), service, civil registration administration, countvectorizer, k-nearest neighbor.*

1. PENDAHULUAN

Pengelolaan Administrasi Kependudukan penting dalam kehidupan di Indonesia karena terkait dengan banyak aktivitas seperti pemilihan umum, pengurusan dokumen kendaraan dan surat tanah, perbankan, dan lain-lain. Kependudukan juga memerlukan Kartu Tanda Penduduk sebagai tanda tempat tinggal bagi penduduk suatu wilayah.

Pemerintah melalui Undang-Undang Republik Indonesia Nomor 23 Tahun 2006 mengeluarkan kebijakan Administrasi Kependudukan yang merujuk pada penataan dan penertiban dokumen dan data kependudukan melalui Pendaftaran Penduduk, Pencatatan Sipil, dan Pengelolaan informasi Administrasi Kependudukan. Hasil administrasi kependudukan ini digunakan untuk pelayanan publik dan pembangunan sektor lainnya.

Dinas Kependudukan dan Pencatatan Sipil bertanggung jawab untuk memberikan pelayanan kepada masyarakat, namun mendapat berbagai tanggapan dari masyarakat. Ada yang positif dan membangun, namun juga ada yang kurang puas. Suku Dinas Kependudukan dan Pencatatan Sipil Kota Jakarta Barat bertugas sesuai dengan peraturan yang berlaku sebagai penyelenggara pemerintahan di bidang administrasi kependudukan di Kota Administrasi Jakarta Barat, yang memiliki jumlah penduduk sebesar 2.434.511 jiwa pada tahun 2020.

Dalam pelaksanaannya, proses administrasi kependudukan di Kota Administrasi Jakarta Barat masih mengalami berbagai kendala dan permasalahan. Berdasarkan pengamatan awal yang telah dilakukan di Suku Dinas Kependudukan dan Pencatatan Sipil Kota Jakarta Barat, masyarakat sering mengeluhkan pelayanan dari petugas yang memberikan arahan tertentu sesuai standar pelayanan namun terkadang masih dianggap di persulit. Hal lain yang sering terjadi dan juga sering menjadi keluhan masyarakat yakni adalah sulitnya atau terbatasnya blanko KTP-el sehingga menjadi kendala dalam waktu penerbitannya sehingga menjadi anggapan bagi masyarakat bahwa mengurus urusan dokumen kependudukan di Dinas Kependudukan dan Pencatatan Sipil masih terus dipersulit.

Untuk meningkatkan pemahaman dan meningkatkan kualitas layanan di Dinas Kependudukan dan Pencatatan Sipil di Kota Administrasi Jakarta Barat, perlu melakukan analisis sentimen terhadap ulasan masyarakat terkait kualitas layanan yang diberikan. Hal ini dilakukan untuk membedakan ulasan positif dan negatif dalam rangka meningkatkan pelayanan. Masyarakat sering mengeluhkan pelayanan petugas, sulitnya blanko KTP-el, dan dianggap dipersulit dalam mengurus dokumen kependudukan.

Analisis sentimen, atau sering disebut juga sebagai *opinion mining*, adalah proses yang digunakan untuk mendeteksi polaritas (positif, negatif, atau netral) dari teks dan mengekstraksi fitur-fitur yang terkait dalam bentuk klasifikasi. Tujuan

utama analisis sentimen adalah untuk memahami dan mengidentifikasi pendapat, opini, atau evaluasi yang terkandung dalam teks. Dalam analisis sentimen, teks umumnya dikategorikan sebagai positif atau negatif berdasarkan emosi, sikap, atau sentimen yang diekspresikan dalam teks tersebut. Hal ini memungkinkan kita untuk mengukur dan memahami pandangan atau respons orang terhadap suatu topik, produk, layanan, atau peristiwa [1]. Analisis sentimen mahasiswa terhadap saran kuisioner tentang penilaian dosen juga pernah dilakukan [2] dengan menggunakan algoritme *K-Nearest Neighbor* dan diperoleh hasil klasifikasi dengan rata-rata akurasi sebesar 81%. Analisis sentimen terhadap ulasan pengguna MRT Jakarta juga pernah dilakukan, diperoleh akurasi sebesar 86% dengan parameter nilai $K=3$ [3]. Pada penelitian [4] dengan analisis sentimen terhadap *review* atau ulasan *Fintech* diperoleh akurasi terbaik sebesar 84,71%. Pernah dilakukan juga penelitian analisis sentimen masyarakat terhadap pembelajaran di era pandemi *Covid-19* oleh [5] yang memperoleh kesimpulan positif dengan akurasi sebesar 71%.

Penelitian ini bertujuan untuk melakukan analisis sentimen masyarakat terhadap Suku Dinas Kependudukan dan Pencatatan Sipil Kota Jakarta Barat. Dalam penelitian ini, digunakan metode analisis sentimen dengan pendekatan machine learning menggunakan fitur kamus sentimen. Ekstraksi fitur dilakukan menggunakan CountVectorizer, sementara algoritme klasifikasi yang digunakan adalah *K-Nearest Neighbor*. Dataset yang digunakan terdiri dari teks yang diperoleh dari hasil survei atau ulasan yang diberikan langsung oleh masyarakat. Dataset ini dapat diakses melalui URL: https://kemsalim.space/ulasan_dukcapil/. Penting untuk dicatat bahwa dataset ini telah dikumpulkan per tanggal 5 September 2022.

2. METODOLOGI PENELITIAN

2.1 Tinjauan Pustaka

2.1.1 Text Mining

Text Mining adalah proses analisis teks yang dilakukan secara otomatis oleh komputer untuk mengekstrak informasi yang terkandung dalam dokumen atau rangkaian teks [6]. Dalam pengertian lain, text mining merupakan teknik dalam penggalian data yang melibatkan teks sebagai sumber datanya, biasanya berupa dokumen, sehingga memungkinkan dilakukannya analisis terhadap hubungan antara dokumen-dokumen tersebut [7].

2.1.2 Ulasan

Ulasan, yang berasal dari kata "ulas" atau dapat disebut juga dengan "kupas", "tafsir", atau "komentar" menurut Kamus Besar Bahasa Indonesia, merujuk pada reaksi atau tanggapan terhadap suatu kejadian. Ulasan yang diberikan pada produk atau aplikasi dianggap penting karena sebagian besar pengguna cenderung melihat ulasan dari pengguna

sebelumnya. Ulasan ini memberikan informasi dan pandangan dari pengguna yang telah menggunakan produk atau aplikasi tersebut, serta memberikan gambaran tentang kualitas, kinerja, atau pengalaman pengguna lainnya. Ulasan ini dapat membantu calon pengguna dalam membuat keputusan atau penilaian sebelum memutuskan untuk menggunakan suatu produk atau aplikasi. Dalam arti lain, ulasan atau komentar juga merupakan suatu teks atau kalimat yang berisi penilaian atau komentar yang bisa dianggap sebagai evaluasi terhadap sebuah ulasan yang biasanya berisi berbagai jenis pengungkapan seperti pujian, pernyataan, interpretasi, maupun komentar lainnya [8].

2.1.3 Analisis Sentimen

Analisis sentimen adalah studi komputerisasi dari opini, perasaan dan emosi yang diungkapkan dalam bentuk teks. *Opinion mining* memiliki tujuan untuk menentukan apakah komentar yang terdapat dalam sekumpulan atau set dokumen teks yang berisi opini tentang suatu objek bersifat positif atau negatif [9]. Dalam pengertian lain, analisis sentimen atau penambahan opini adalah bidang studi yang berkaitan dengan menganalisis pendapat orang tentang berbagai entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, dan topik. Fokus utama dari analisis sentimen ini adalah pada pendapat pribadi yang mengekspresikan atau menyiratkan sentimen positif atau negatif. Sebagian besar analisis sentimen ini terkait dengan pendapat yang diungkapkan oleh orang-orang melalui media sosial [10].

2.1.4 Pelayanan Administrasi Kependudukan

Menurut Undang-Undang Nomor 24 Tahun 2013 tentang Administrasi Kependudukan, Administrasi Kependudukan mencakup serangkaian kegiatan untuk mengatur dan memperbarui dokumen dan data kependudukan melalui Pendaftaran Penduduk, Pencatatan Sipil, Pengelolaan Informasi Administrasi Kependudukan, dan penggunaan hasilnya untuk pelayanan publik. Dokumen kependudukan merujuk pada dokumen resmi yang dikeluarkan oleh lembaga pelaksana yang memiliki kekuatan hukum dan digunakan sebagai bukti otentik yang dihasilkan dari layanan Pendaftaran Penduduk dan Pencatatan Sipil. Kegiatan Pelayanan Administrasi Kependudukan terdiri dari Pendaftaran Penduduk dan Pencatatan Sipil, yang masing-masing berfokus pada penerbitan dokumen kependudukan dan pencatatan peristiwa penting dalam registri Pencatatan Sipil.

2.1.5 CountVectorizer

Mesin tidak dapat memahami karakter dan kata-kata. Jadi ketika berhadapan dengan data teks kita perlu merepresentasikannya ke dalam bentuk angka agar dapat dipahami oleh mesin. *CountVectorizer* merupakan sebuah metode atau proses pengolahan

dokumen atau teks yang mengubah teks menjadi representasi vektor. *CountVectorizer* digunakan untuk menghitung frekuensi kata yang muncul dalam dokumen atau kalimat, dan kemudian mengubahnya menjadi vektor berdasarkan frekuensi tersebut [11]. Adapun contoh cara kerja *CountVectorizer* pada Gambar 1

text = ['Hello my name is kemil, this is my python notebook']								
	hello	is	kemil	my	name	notebook	python	this
0	1	2	1	2	1	1	1	1

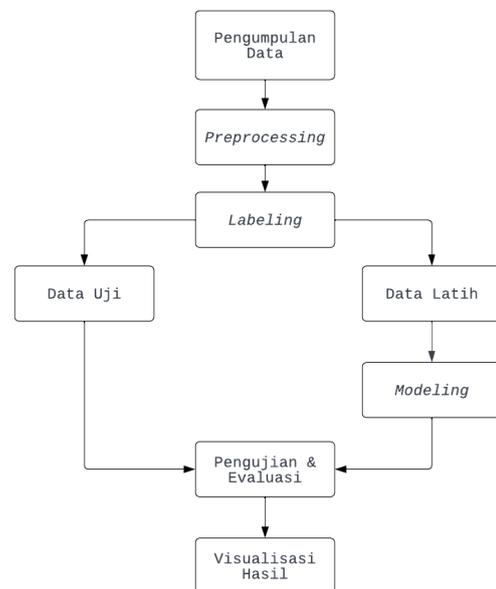
Gambar 1. Cara Kerja *CountVectorizer*

2.2 Data Penelitian

Dataset yang digunakan dalam penelitian ini diperoleh dari formulir ulasan yang diisi oleh responden atau pengguna yakni masyarakat sendiri yang telah menerima layanan di kantor Suku Dinas Kependudukan dan Pencatatan Sipil Kota Administrasi Jakarta Barat. Data tersebut dikumpulkan melalui *url* : https://kemsalim.space/ulasan_dukcapil/.

2.3 Penerapan Metode

Dalam membangun aplikasi analisis sentimen yang diimplementasikan dalam penelitian ini, dilakukan beberapa langkah. Langkah-langkah tersebut mewakili setiap proses dan desain penelitian ini dari awal hingga akhir aplikasi bekerja. Tahapan yang dilakukan dapat dilihat pada Gambar 2.



Gambar 2. Tahapan Metode

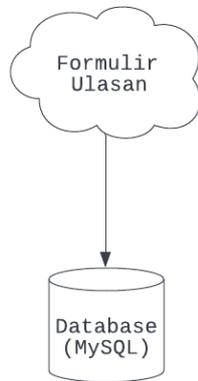
2.3.1 Pengumpulan Data

Pada tahapan pengumpulan data dilakukan dan dikumpulkan melalui laman: https://kemsalim.space/ulasan_dukcapil/ yang disebar melalui barcode url yang ditempel dimeja

layanan, ulasan yang diisi pada laman tersebut diisi langsung oleh masyarakat yang telah menerima layanan di Kantor Suku Dinas Kependudukan dan Pencatatan Sipil Kota Administrasi Jakarta Barat yang kemudian masuk ke dalam basis data (database) MySQL. Pada Gambar 3 Barcode Formulir Ulasan merupakan media utama yang digunakan untuk membantu mengumpulkan data atau ulasan dari masyarakat langsung. Barcode tersebut dicetak dalam bentuk stiker yang dicetak lalu ditempel di meja pelayanan, untuk mempermudah masyarakat ketika berkenan untuk memberi ulasan terkait layanan yang diterima.



Gambar 3. Barcode Formulir Ulasan



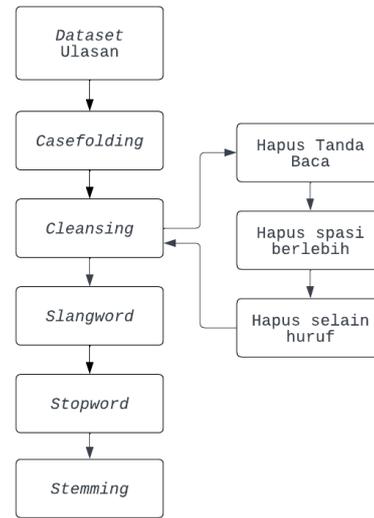
Gambar 4. Tahapan Pengumpulan Data

Pada gambar 3, merupakan sebuah gambar yang digunakan sebagai sarana untuk memudahkan warga atau masyarakat dalam memberikan ulasan terhadap layanan. Tahap selanjutnya digambarkan pada gambar 4, di mana ulasan yang diberikan oleh masyarakat melalui tautan langsung yang terdapat pada barcode, akan langsung disimpan ke dalam basis data yang telah disiapkan.

2.3.2 Preprocessing

Dalam tahap *preprocessing*, dilakukan penyaringan, penghilangan, dan perbaikan kata pada data ulasan melalui beberapa proses seperti *case folding*, *cleansing*, perubahan *slang word*,

penghapusan *stop word*, dan *stemming*. Hal ini bertujuan untuk menghasilkan data ulasan yang lebih terstruktur dan jelas, yang biasa disebut sebagai *clean text*. Ilustrasi proses tahap *preprocessing* data dapat dilihat pada gambar 5 dan hasil *preprocessing* dengan salah satu ulasan yang diterima pada tabel 1 sebagai berikut :



Gambar 5. Tahapan Preprocessing

Tabel 1. Contoh Penerapan Proses Processing Data

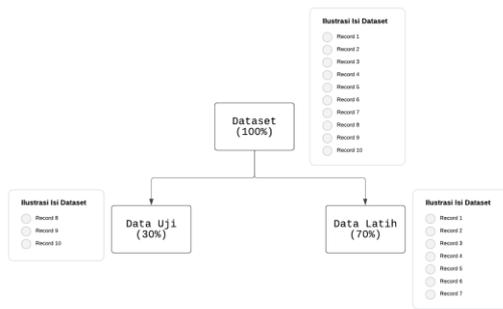
Tahapan	Ulasan
Teks Awal	Terima kasih untuk bantuannya Kantor Dukcapil Kelurahan Kalideres, telah membantu perekaman EKTP Lansia untuk Bapak saya yang sudah lansia dan kondisi sakit. Terima kasih banyak.
Case Folding	terima kasih untuk bantuannya kantor dukcapil kelurahan kalideres, telah membantu perekaman ektp lansia untuk bapak saya yang sudah lansia dan kondisi sakit. terima kasih banyak.
Cleansing	terima kasih untuk bantuannya kantor dukcapil kelurahan kalideres telah membantu perekaman ektp lansia untuk bapak saya yang sudah lansia dan kondisi sakit terima kasih banyak
Slang Word	terima kasih untuk bantuannya kantor dukcapil kelurahan kalideres telah membantu perekaman ktp lanjut usia untuk bapak saya yang sudah lanjut usia dan kondisi sakit terima kasih banyak
Stop Word	terima kasih bantuannya kantor dukcapil kelurahan kalideres membantu perekaman ktp lanjut usia lanjut usia kondisi sakit terima kasih
Stemming	terima kasih bantu kantor dukcapil lurah kalideres bantu rekam ktp lanjut usia lanjut usia kondisi sakit terima kasih

2.3.3 Labeling

Pada tahapan labeling dilakukan penandaan label (kelas) berdasarkan fitur atau karakteristik yang terdapat pada sebuah dokumen atau kalimat [12] [13]. Pada fase ini teks ulasan yang sudah melalui tahapan proses preprocessing dan menghasilkan teks bersih akan diberikan kategori positif atau negatif. Kelas positif dimaksudkan untuk ulasan yang mengandung pernyataan yang baik, berisi dukungan. Sedangkan kelas negatif cenderung untuk ulasan yang mengkoreksi, memberi komentar atau menunjukkan rasa kecewa terhadap pelayanan yang diterima.

2.3.4 Pembagian Data

Pada langkah ini, ulasan yang telah diberi label atau klasifikasi kemudian dipisahkan menjadi dua kelompok, yaitu data uji dan data latih. Proses pemisahan data dilakukan dengan membagi dataset menjadi dua bagian, di mana sekitar 70% digunakan sebagai data latih (training data) dan sekitar 30% digunakan sebagai data uji (test data). Ilustrasi tahapan pembagian data dapat dilihat pada gambar 6 berikut :



Gambar 6. Tahapan Pembagian Data

2.3.5 Modeling



Gambar 7. Tahapan Modeling

Pada gambar 7 dapat diketahui bahwa tahapan pemodelan ini dilakukan untuk mengekstraksi teks ulasan data latih yang diubah menjadi representasi vektor menggunakan teknik

CountVectorizer. Dalam fase ini, terdapat lima proses utama yang dilakukan dalam pembuatan model latih, yaitu seleksi data latih, pembuatan list kata, pencarian fitur kata, pembuatan vektor kosong, dan pembuatan representasi vektor.

2.3.6 Klasifikasi K-Nearest Neighbor

Pada tahapan ini dilakukan klasifikasi menggunakan algoritme K-Nearest Neighbor (KNN) yaitu langkah yang dapat dilakukan setelah terdapat satu atau lebih kumpulan model data latih. Model latih tersebut merupakan data latih yang telah melalui tahap *modeling*. Untuk menerapkan model klasifikasi menggunakan KNN, terdapat empat (4) buah proses utama yang dilakukan yaitu : membuat representasi vektor uji, menghitung jarak antar data, mencari dan menemukan terdekat berdasarkan nilai-K, dan hitung nilai probabilitas label sentimen. Ilustrasi tahapan klasifikasi dapat dilihat pada gambar 8 dibawah dan dengan persamaan *K-Nearest Neighbor* berikut:

$$d_{(x,y)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{1}$$



Gambar 8. Tahapan Klasifikasi

2.3.7 Rancangan Pengujian

Pada tahap pengujian ini, dilakukan evaluasi terhadap model latih menggunakan algoritme *K-Nearest Neighbor* (KNN) untuk menentukan tingkat akurasi, presisi, dan *recall*. Evaluasi dilakukan dengan membandingkan data hasil prediksi dari model (hasil fase klasifikasi) dengan data aktual (data yang telah dilabeli). Data hasil prediksi merupakan sekumpulan data yang diperoleh dari pemrosesan dataset menggunakan algoritme *K-Nearest Neighbor* (KNN) dengan beberapa nilai K, yaitu: K=3, K=5, K=7, K=9, dan K=11.

a. Akurasi

Akurasi merupakan ukuran yang menggambarkan tingkat kesesuaian atau kedekatan antara nilai prediksi yang diberikan oleh model

dengan nilai sebenarnya atau *ground truth*., dengan persamaan sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

b. Presisi

Presisi adalah ukuran yang menggambarkan tingkat ketepatan atau keakuratan sistem dalam memberikan jawaban yang sesuai atau relevan dengan data yang diminta dalam konteks persamaan atau perbandingan, dengan persamaan berikut :

$$Presisi = \frac{TP}{TP+FP} \quad (3)$$

c. Recall

Recall adalah ukuran yang menggambarkan tingkat keberhasilan suatu sistem dalam menemukan kembali atau mengingat kembali sebuah informasi yang relevan atau penting dalam konteks yang diberikan, dengan persamaan sebagai berikut:

$$Recall = \frac{TP}{(TP+FN)} \quad (4)$$

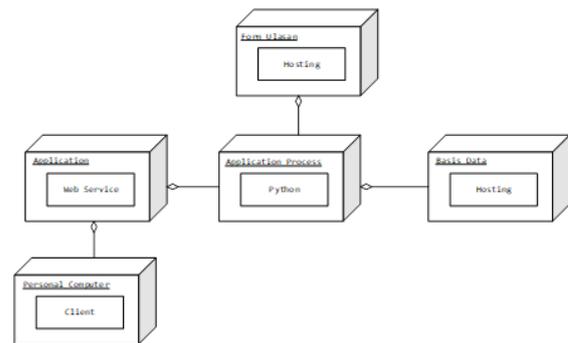
Pengukuran tingkat akurasi, presisi, dan *recall* dapat diketahui melalui *confusion matrix* dengan persamaan sebagai berikut :

- True Positive* (TP) adalah data yang positif dan berhasil diprediksi dengan benar. Contohnya, jika terdapat sebuah teks ulasan yang memiliki label positif, dan model latihan yang dibuat berhasil memprediksi teks ulasan tersebut juga sebagai ulasan positif..
- True Negative* (TN) adalah data yang negatif dan berhasil diprediksi dengan benar. Contohnya, jika terdapat sebuah teks ulasan yang memiliki label negatif, dan model latihan yang dibuat berhasil memprediksi teks ulasan tersebut juga sebagai ulasan negatif.
- False Positive* (FP) adalah data yang sebenarnya negatif, tetapi diprediksi sebagai positif. Misalnya, jika terdapat sebuah teks ulasan yang memiliki label negatif, tetapi model latihan yang dibuat memprediksi teks ulasan tersebut sebagai ulasan positif.
- False Negative* (FN) merupakan adalah data yang sebenarnya positif, tetapi diprediksi sebagai negatif. Misalnya, jika terdapat sebuah teks ulasan yang memiliki label positif, tetapi model latihan yang dibuat memprediksi teks ulasan tersebut sebagai ulasan negatif.

3. HASIL DAN PEMBAHASAN

3.1 Prototipe Sistem Analisis Sentimen

Dalam penelitian ini, dilakukan pengembangan prototipe sistem analisis pendapat masyarakat terhadap layanan Disdukcapil dengan menerapkan metode klasifikasi K-Nearest Neighbor. Tujuan utama dari pengembangan prototipe ini adalah untuk mempermudah analisis pendapat bagi pengguna, sementara juga menguji keefektifan metode dan algoritma yang diusulkan dalam melakukan klasifikasi pendapat dengan baik. Dalam penelitian ini, prototipe dibuat menggunakan bahasa pemrograman Python dan beberapa perpustakaan pendukung yang relevan.



Gambar 9. Deployment Diagram Prototipe Sistem

Diagram *deployment* pada Gambar 9 menunjukkan rancangan sistem prototipe analisis sentimen yang dikembangkan dalam penelitian ini. Sistem analisis sentimen ini menggunakan layanan *hosting* untuk menyimpan basis data dan formulir layanan *online*-nya, serta menggunakan *framework Flask* untuk menghosting aplikasinya secara lokal.



Gambar 10. Landing Page Sistem Aplikasi

3.2 Pengujian dan Evaluasi

Pada penelitian ini, dilakukan pengujian dari sisi akurasi, presisi dan *recall* pada implementasi algoritme *K-Nearest Neighbor* (KNN) dalam memprediksi label untuk data uji. Selain pada sisi akurasi, presisi, dan *recall* pengujian ini peneliti juga menguji nilai K yang variatif berdasarkan yang telah di tetapkan yaitu K=3, K=5, K=7, K=9, dan K=11.

Hasil sampel prediksi oleh algoritme KNN dengan nilai K=3 dapat dilihat pada Tabel 2.

Tabel 2. Sampel Data Hasil Prediksi

No	Ulasan	Label Aktual	Label Prediksi
1	CEPAT PELAYANNYA MUDAH PROSESNYA.	Positif	Positif
2	cukup mudah sih prosesnya, tapi sayangnya masih harus bulak balik ke kelurahan.	Positif	Positif
3	Prosesnya cepat, tapi kenapa tidak bisa langsung jadi EKTP nya?	Negatif	Positif
...
83	Kurang diperhatikan lagi kendala yang dialami, jadi saya merasa kurang mengerti. Bertele-tele penjelasannya.	Negatif	Negatif

Pada Tabel 3 disajikan sampel data hasil prediksi, kolom label aktual merupakan data label yang diperoleh melalui proses labeling. Sedangkan label prediksi merupakan hasil data label dari proses klasifikasi menggunakan KNN. Keseluruhan hasil prediksi data uji sebanyak 83 data kemudian di representasikan ke dalam bentuk *confusion matrix*. *Confusion matrix* yang terbentuk dari nilai K=3 dapat dilihat pada Tabel 3.

Tabel 3. *Confusion Matrix* Pengujian K=3

		Nilai Aktual	
		Positif	Negatif
Nilai Prediksi	Positif	92	20
	Negatif	0	3

Berdasarkan tabel 3, maka perolehan nilai akurasi, presisi dan *recall* menggunakan rumus yang telah dijabarkan dalam persamaan dapat dilihat pada Tabel 4.

Tabel 4. Nilai Pengujian K=3

	Pengujian	
	Positif	Negatif
Akurasi	$\frac{92 + 3}{92 + 3 + 20 + 0}$	83%
Presisi	$\frac{92}{92 + 20}$	82%
Recall	$\frac{92}{92 + 0}$	100%

Pengujian di atas dilakukan secara bertahap dan berulang dengan variasi nilai K yang berbeda-beda. Sehingga dapat diketahui hasil pengujian secara adalah seperti pada Tabel 5.

Tabel 5. Hasil Pengujian

	Akurasi	Presisi	Recall
K=3	83%	82%	100%
K=5	82%	81%	100%
K=7	81%	81%	100%
K=9	81%	81%	100%
K=11	81%	81%	100%

Berdasarkan Tabel 5, hasil pengujian dapat diketahui dan disimpulkan bahwa hasil pengujian menunjukkan algoritme KNN mampu memperoleh nilai pengujian tertinggi menggunakan nilai K=3 dengan akurasi 83%, presisi 82% dan recall 100%.

4. KESIMPULAN

Berdasarkan 386 ulasan, arah pandang atau sentimen masyarakat yang menerima layanan di Kantor Suku Dinas Kependudukan dan Pencatatan Sipil Kota Administrasi Jakarta Barat positif 311 ulasan sebesar (80,57%) dan 75 ulasan negatif (19,43%) sejak 5 September 2022 sampai dengan 15 Desember 2022.

Terdapat beberapa tahap utama dalam penelitian ini, yaitu pengumpulan data, *preprocessing*, *labeling*, *modeling*, dan klasifikasi menggunakan algoritme *K-Nearest Neighbor* (KNN). Tahap *preprocessing* yang dilakukan dengan baik memiliki peran penting dalam menghasilkan hasil yang optimal pada tahap selanjutnya.

Dalam penelitian ini, teknik ekstraksi fitur *CountVectorizer* dan algoritme *K-Nearest Neighbor* (KNN) telah terbukti efektif dalam melakukan analisis sentimen, dengan hasil pengujian dan evaluasi yang sangat baik, yaitu akurasi sebesar 83%, presisi sebesar 82%, dan *recall* sebesar 100% dengan menggunakan nilai K=3.

DAFTAR PUSTAKA

- [1] A. Alsaedi and M. Z. Khan, "A study on sentiment analysis techniques of Twitter data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 2, pp. 361–374, 2019, doi: 10.14569/ijacsa.2019.0100248.
- [2] N. S. N. Salam, A. A. Supianto, and A. R. Perdanakusuma, "Analisis Sentimen Opini Mahasiswa Terhadap Saran Kuesioner Penilaian Kinerja Dosen dengan Menggunakan TF-IDF dan K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 6, pp. 6148–6156, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/5649>
- [3] A. A. Paramitha, Indriati, and Y. A. Sari, "Analisis Sentimen Terhadap Ulasan Pengguna MRT Jakarta Menggunakan Information Gain dan Modified K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 4, pp. 1125–1132, 2020.
- [4] S. Surohman, S. Aji, R. Rousyati, and F. F. Wati, "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K- Nearest Neighbor," *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 93–105, 2020, doi: 10.31294/evolusi.v8i1.7535.
- [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,” *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, no. September, pp. 311–319, 2021.
- [6] H. Asril, Mustakim, and I. Kamila, “Klasifikasi Dokumen Tugas Akhir Berbasis Text Mining menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor,” *Semin. Nas. Teknol. Inf. Komun. dan Ind.*, vol. 0, no. 0, pp. 2579–5406, 2019, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/7995>
- [7] P. P. A. Indriya Dewi Onantya, Indriati, “Analisis Sentimen Pada Ulasan Aplikasi BCA Mobile Menggunakan BM25 Dan Improved K-Nearest Neighbor,” *J-Ptiik.Ub.Ac.Id*, vol. 3, no. 3, pp. 2575–2580, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4754>
- [8] Y. M. Febrianti, I. Indriati, and A. W. Widodo, “Analisis Sentimen Pada Ulasan ‘Lazada ’ Berbahasa Indonesia Menggunakan K-Nearest Neighbor (K-NN) Dengan Perbaikan Kata Menggunakan Jaro Winkler Distance,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3689–3698, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [9] R. R. A. Siregar, Z. U. Siregar, and R. Arianto, “Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor,” *Kilat*, vol. 8, no. 1, 2019, doi: 10.33322/kilat.v8i1.421.
- [10] J. A. Septian, T. M. Fahrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor,” *J. Intell. Syst. Comput.*, pp. 43–49, 2019, [Online]. Available: <https://t.co/9W1oaWpfD5>
- [11] S. and Munawar, “Sistem Pendeteksi Berita Hoax Di Media Sosial Dengan Teknik Data Mining Scikit Learn,” *J. Ilmu Komput. Vol. 4 Nomor 2 Desember 2019*, vol. 4, pp. 173–179, 2019.
- [12] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, “Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM),” vol. 9, pp. 376–390, 2020.
- [13] R. Mahendrajaya, G. A. Buntoro, M. B. Setyawan, F. Teknik, and U. M. Ponorogo, “Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine,” 2019