



Implementasi Optimasi NLP dan KNN untuk User Review Aplikasi SAMPEAN Cirebon

Mesi Febima¹, Lena Magdalena², Marsani Asfi³, Muhammad Hatta⁴, Rifqi Fahrudin⁵

^{1,2,3,4,5}Sistem Informasi, Fakultas Teknologi Informasi, Universitas Catur Insan Cendekia

mesi.febima@cic.ac.id, lena.magdalena@cic.ac.id, marsani.asfi@cic.ac.id, muhammad.hatta@cic.ac.id,

rifqi.fahrudin@cic.ac.id

Abstract

The use of information technology in the personnel administration process plays an active role in improving public services for civil servants (ASN) in the Cirebon City Government by providing accurate data for decision-making. One of the smart city applications that assists ASN in Cirebon City in supporting personnel administration activities is the SAMPEAN Cirebon City application. However, to ensure that this application is truly effective and meets user needs, it is important to analyze user reviews provided through application reviews. One effective method for analyzing user reviews is by using Natural Language Processing (NLP) and machine learning techniques. The NLP technique and classification model used is the KNN algorithm. The purpose of this research is to provide valuable input for application developers in improving the quality and performance of the SAMPEAN application. The research results show that by testing accuracy using the confusion matrix with K values of 3, 5, 7, and 9, it was found that K=9 provides the best performance with a balance between precision, recall, F1-Score, and accuracy. The model achieved a precision of 64%, recall of 90%, F1-Score of 75%, and accuracy of 62%. It can be concluded that with the optimization of the K parameter in KNN, the higher the K value, the higher the accuracy. This emphasizes the importance of selecting the right parameters to enhance the effectiveness of machine learning models in various Natural Language Processing (NLP) applications.

Keywords: NLP, KNN, User Reviews, Text Classification, Optimization.

Abstrak

Penggunaan teknologi informasi dalam proses administrasi kepegawaian sangat berperan aktif dalam peningkatan pelayanan publik bagi ASN Pemda Kota Cirebon untuk penyajian data yang akurat dalam rangka pengambilan keputusan. Salah satu aplikasi *smart city* yang membantu ASN Kota Cirebon dalam menunjang kegiatan administrasi kepegawaian adalah aplikasi SAMPEAN Kota Cirebon. Namun, untuk memastikan bahwa aplikasi ini benar-benar efektif dan memenuhi kebutuhan pengguna, penting untuk menganalisis ulasan pengguna (*user reviews*) yang diberikan oleh pengguna melalui ulasan aplikasi. Salah satu metode yang efektif untuk menganalisis ulasan pengguna adalah dengan menggunakan teknik Natural Language Processing (NLP) dan machine learning. Teknik NLP dan model klasifikasi yang digunakan adalah algoritma KNN. Tujuan penelitian ini untuk memberikan masukan berharga bagi pengembang aplikasi dalam meningkatkan kualitas dan kinerja aplikasi SAMPEAN. Hasil penelitian menunjukkan bahwa dengan pengujian akurasi menggunakan confusion matrix dengan nilai K=3,5,7,9 didapatkan hasil bahwa nilai K=9 memberikan performa terbaik dengan keseimbangan antara precision, recall, F1-Score, dan akurasi. Dengan model mencapai precision sebesar 64%, recall 90%, F1-Score 75%, dan akurasi 62%. Dapat disimpulkan dengan Optimalisasi parameter K dalam KNN dengan semakin tinggi nilai K maka semakin tinggi tingkat akurasinya. Hal ini menegaskan pentingnya pemilihan parameter yang tepat untuk meningkatkan efektivitas model machine learning dalam berbagai aplikasi Natural Language Processing (NLP).

Kata kunci: NLP, KNN, User Reviews, Klasifikasi Teks, Optimasi.

1. Pendahuluan

Di era digital saat ini, penggunaan teknologi informasi dalam proses administrasi kepegawaian sangat berperan aktif dalam peningkatan pelayanan publik bagi ASN

Pemda Kota Cirebon untuk penyajian data yang akurat dalam rangka pengambilan keputusan maupun kebijakan manajemen kepegawaian. Sejumlah aplikasi *smart city* yang membantu ASN Kota Cirebon dalam menunjang kegiatan sehari-hari diantaranya Cirebon Wistakon atau

wisata Kota Cirebon, Cirebon Lengko (layanan elektronik kesehatan online), Cirebon Brojol Aja Klalen (Akte Langsung Jadi Kalau Lapor Lewat Online) dengan konsep one day service, Cirebon Sedulur (Sistem Elektronik Pelayanan Digital Kelurahan), Cirebon Segi Jamblang (sistem elektronik kepegawaian dan evaluasi kinerja secara gamblang), dan SAMPEAN (Sistem Aplikasi Mobile Kepegawaian).

Pemda Kota Cirebon melalui penetapan kebijakan Perwal Kota Cirebon Nomor 52 Tahun 2022 tentang Portal Layanan Berbasis Elektronik Terintegrasi di Lingkungan Pemerintah Daerah Kota Cirebon dan Kepwal Kota Cirebon Nomor 100/Kep.404- DKIS/2022 tentang Penetapan Sistem Administrasi Manajemen Pemerintahan Sebagai Portal Layanan Administrasi Pemerintahan Berbasis Elektronik Terintegrasi telah menetapkan dan meluncurkan kepada publik Aplikasi Sistem Administrasi Manajemen Pemerintahan yang selanjutnya disebut SAMPEAN sebagai implementasi dari Penyelenggaraan kebijakan dan Manajemen ASN [1].

Aplikasi SAMPEAN (Sistem Aplikasi Mobile Kepegawaian) merupakan aplikasi berbasis mobile yang dapat digunakan oleh seluruh aparatur di Pemerintah Daerah Kota Cirebon. Implementasi aplikasi SAMPEAN sebagai Aplikasi layanan Administrasi Pemerintahan di Kota Cirebon diharapkan dapat mengintegrasikan data, informasi, serta layanan dalam administrasi pemerintahan yang berbasis elektronik [1]. Aplikasi SAMPEAN yang digunakan oleh ASN Kota Cirebon merupakan salah satu contoh aplikasi yang memberikan berbagai layanan kepada penggunanya. Aplikasi ini dirancang untuk memudahkan ASN dalam mengakses berbagai informasi bagi pegawai, memonitor golongan/pangkat kepegawaian, input tugas-tugas dalam laporan kinerja harian, serta kehadiran yang berupa pengecekan hak (gaji) dan kewajiban kedatangan kantor pegawai. Namun, untuk memastikan bahwa aplikasi ini benar-benar efektif dan memenuhi kebutuhan pengguna, penting untuk menganalisis ulasan pengguna (*user reviews*) yang diberikan oleh pengguna melalui ulasan aplikasi.

Ulasan pengguna (*user reviews*) mengandung informasi berharga mengenai kepuasan, masalah, dan saran yang dapat digunakan untuk meningkatkan kualitas aplikasi SAMPEAN Kota Cirebon. *Google Play Store* merupakan *platform* utama di mana pengguna dapat memberikan *user review* mengenai aplikasi SAMPEAN Kota Cirebon. Banyaknya ulasan yang masuk setiap hari, analisis manual menjadi tidak praktis dan memakan waktu. Oleh karena itu, untuk mengekstraksi informasi berguna dari sejumlah besar *user reviews*, diperlukan teknik analisis yang efisien dan akurat. Salah satu metode yang efektif untuk menganalisis ulasan pengguna adalah dengan menggunakan teknik Natural Language Processing (NLP) dan machine learning. Natural Language Processing (NLP) adalah cabang ilmu komputer yang berkaitan dengan pemahaman dan

pemrosesan bahasa manusia secara alami oleh komputer [2]. NLP memungkinkan komputer untuk memahami, menafsirkan, dan memanipulasi bahasa manusia, sedangkan *machine learning* memungkinkan sistem untuk belajar dari data dan meningkatkan kinerjanya dari waktu ke waktu.

Optimasi algoritma *machine learning* dalam NLP bertujuan untuk meningkatkan akurasi dan efisiensi analisis sentimen dari *user review*. Tahapan NLP untuk memproses dan menganalisis mencakup pengumpulan data ulasan pengguna, preprocessing data, pemilihan dan penerapan algoritma *machine learning* serta evaluasi kinerja model yang dihasilkan. Analisis sentimen adalah proses yang menggunakan Natural Language Processing (NLP) untuk secara otomatis mengumpulkan sikap, pendapat, pandangan, dan emosi dari teks, audio, kicauan, dan sumber data[3]. Analisis Sentimen dapat digunakan sebagai alat untuk penelitian terkait, karena dapat membantu mengidentifikasi dan menganalisis perasaan dan opini yang terkait dengan topik tersebut di media sosial [4]. Analisis sentimen dalam ulasan Aplikasi SAMPEAN tidak hanya memberikan wawasan tentang kepuasan dan ketidakpuasan pengguna, tetapi juga menawarkan sudut pandang yang unik untuk memahami preferensi serta keefektifan dari penggunaan Aplikasi SAMPEAN bagi seluruh aparatur di Pemerintah Daerah Kota Cirebon.

Klasifikasi adalah bentuk dasar dari analisa data dan sering juga dikenal sebagai teknik untuk menentukan kategori kelas dari data-data yang sudah ada [5]. Klasifikasi termasuk ke dalam algoritma *supervised learning*. Pendekatan model klasifikasi dalam analisis sentimen adalah salah satu algoritma yang efektif untuk mengkategorikan ulasan pengguna ke dalam klasifikasi ulasan teknis, dan ulasan pengalaman pengguna. Ulasan teknis berfokus pada aspek teknik aplikasi, seperti kinerja, bug, atau masalah teknis lainnya dan ulasan pengalaman pengguna berfokus pada ulasan yang menceritakan pengalaman spesifik pengguna dengan aplikasi, yang mencakup cerita atau situasi.

Dalam penelitian ini mengimplementasikan teknik NLP dan model klasifikasi untuk menganalisis sentimen dari *user reviews* aplikasi SAMPEAN di Kota Cirebon. Dengan menggunakan algoritma K-Nearest Neighbors (KNN) sebagai model klasifikasi. Algoritma K-Nearest Neighbor atau KNN merupakan algoritma dalam melakukan klasifikasi objek berdasarkan jarak data pembelajaran baru berdasarkan K tetangga yang paling dekat[6]. Tahapannya dalam memprediksi menggunakan algoritma KNN adalah melakukan data training dan data uji berdasarkan kategori pada sampel atau pada data yang lampau dan sesuai dengan k sampel training di mana tetangga terdekat dengan sampel uji, selanjutnya dimasukkan ke dalam kategori yang memiliki probabilitas terbesar [7]. Algoritma KNN diharapkan mampu memberikan hasil analisis yang akurat dan membantu pihak pengembang aplikasi untuk memahami sentimen pengguna dengan lebih baik.

Penelitian terkait analisis sentimen ulasan aplikasi telah dilakukan sebelumnya. Beberapa penelitian terdahulu yang berkaitan dengan analisis sentimen adalah penelitian yang dilakukan oleh [8] tentang klasifikasi sentimen publik terhadap jenis vaksin covid-19 menggunakan NLP dan KNN. Berdasarkan hasil penelitian yang dilakukan, sentimen positif tertinggi pada jenis vaksin Pfizer disusul Moderna yakni 47,30% dan 46,20%. Sementara vaksin AstraZeneca mendapat peringkat sentimen terendah dari ketiganya yakni 40,09%. Kesimpulan dari penelitian ini adalah masyarakat lebih percaya pada jenis vaksin Pfizer dibandingkan dengan jenis vaksin lainnya. Pada penelitian yang dilakukan [9] dalam penelitiannya tentang analisis sentimen *E-commerce* Shopee menggunakan algoritma KNN. Dari hasil penelitian, dapat disimpulkan bahwa model Sentiment Analysis yang dikembangkan menggunakan algoritma NLP dan KNN menunjukkan kinerja yang memuaskan. Akurasi model mencapai 82%, dan nilai dari cross-validation score sebesar 80%. Evaluasi lainnya juga memberikan hasil yang positif, menandakan bahwa model ini efektif dalam melakukan analisis sentimen pada dataset ulasan pengguna aplikasi Shopee.

Selanjutnya penelitian yang dilakukan oleh [10] tentang implementasi algoritma KNN untuk analisis sentimen pada pengguna FLIP, dari hasil penelitiannya menunjukkan 77,67% dari data uji benar terklasifikasi ke dalam kelas ulasan positif dengan nilai precision dan recall tinggi yaitu 82,67% dan 86,92%. Selain itu, dari hasil penerapan metode dalam mengklasifikasikan data ulasan pengguna Flip dengan perbandingan data latih dan data uji sebesar 80%:20% diperoleh tingkat akurasi klasifikasi dengan algoritma K-Nearest Neighbor sebesar 76,68%.

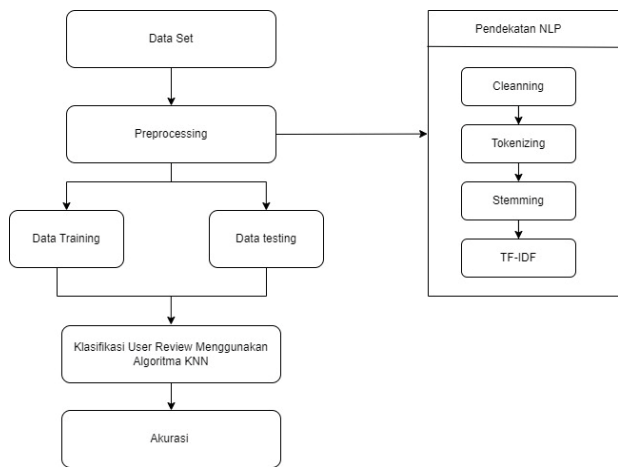
Dalam penelitian ini, fokus utama adalah pada optimasi *machine learning* untuk NLP dan algoritma klasifikasi KNN dalam analisis sentimen ulasan pengguna aplikasi SAMPEAN di Kota Cirebon. Dengan mengoptimalkan teknik ini, diharapkan dapat diperoleh hasil analisis yang lebih akurat dan relevan, sehingga dapat memberikan masukan berharga bagi pengembang aplikasi untuk meningkatkan kualitas dan kinerja aplikasi SAMPEAN. Data yang digunakan dalam penelitian ini adalah data *reviews* pada aplikasi SAMPEAN yang diperoleh pada *Play Store*. Data *reviews* pengguna tersebut dilakukan teknik *preprocessing* menggunakan NLP. Selanjutnya data teks diproses ditentukan polaritas dan subjektivitas untuk klasifikasi sentimen, hal ini dikarenakan algoritma *machine learning* tidak dapat mengklasifikasi data teks [5]. Selanjutnya proses klasifikasi dengan algoritma KNN digunakan untuk pemodelan klasifikasi menjadi 2 yaitu ulasan teknis, dan ulasan pengalaman pengguna.

Dengan demikian, implementasi *machine learning* pada NLP dan model klasifikasi dalam analisis sentimen ini diharapkan dapat memberikan kontribusi yang signifikan dalam upaya meningkatkan pengalaman

pengguna dan efektivitas layanan aplikasi SAMPEAN di Kota Cirebon. Hasil dari penelitian ini dapat membantu pengembang atau *developer* aplikasi SAMPEAN agar dapat meningkatkan kualitas aplikasi berdasarkan klasifikasi sentimen ulasan teknis dan ulasan pengalaman pengguna dari *user reviews* aplikasi SAMPEAN Kota Cirebon.

2. Metode Penelitian

Pada penelitian ini, alur penelitian dibagi menjadi 4 tahapan yang dimulai dari data set yang diambil dari *user review* pada *Play Store*. Selanjutnya data set akan dilakukan *preprocessing data* menggunakan pendekatan NLP. Selanjutnya data set dibagi menjadi data *training* dan data *testing* yang digunakan pada proses klasifikasi *user review* dengan 2 klasifikasi sentimen ulasan teknis, dan sentimen ulasan pengalaman pengguna. Ulasan teknis berfokus pada aspek teknik aplikasi, seperti kinerja, bug, atau masalah teknis lainnya dan ulasan pengalaman pengguna berfokus pada ulasan yang menceritakan pengalaman spesifik pengguna dengan aplikasi, yang mencakup cerita atau situasi. Proses klasifikasi menggunakan algoritma KNN. Tahapan akhir dari penelitian ini adalah melakukan uji akurasi dari hasil klasifikasi menggunakan algoritma KNN. Adapun tahapan yang dimaksud dapat dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

2.1. Data Set

Penelitian ini menggunakan teknik *crawling* untuk mengumpulkan data set *user reviews* pada aplikasi SAMPEAN yang diperoleh pada link '*dkis.egov.sampean.mobile*' *Play Store*. Data Set yang diperoleh sejumlah 77 data *user reviews*.

2.2. Preprocessing

Preprocessing data merupakan langkah penting dalam penelitian yang bertujuan untuk membersihkan, merapikan, dan mengubah data mentah ke dalam format yang lebih sesuai untuk analisis lebih lanjut.

Preprocessing data dimaksudkan untuk membersihkan data teks dengan cara yang relevan untuk dianalisis kedepannya. Penelitian ini tahapan Pra-Pemrosesan Data dilakukan dengan pendekatan pendekatan NLP (*Natural Language Processing*).

2.3. Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah cabang ilmu komputer yang berkaitan dengan pemahaman dan pemrosesan bahasa manusia secara alami oleh komputer. NLP berfokus pada pengembangan metode dan algoritma untuk mengenali, memahami, menganalisis, dan memanipulasi teks dan bahasa manusia dalam bentuk yang dapat dipahami oleh mesin [2]. Pendekatan NLP terdiri dari beberapa tahapan yaitu:

Cleansing, Pertama pada proses cleansing, karakter yang tidak penting seperti tanda baca, emoji, URL, hashtag, dan angka dihapus untuk menghapus data dari item yang tidak relevan. Hal ini membuat data lebih bersih dan lebih mudah diproses pada langkah selanjutnya. Tujuan dari tahap cleansing ini yakni untuk membersihkan data dari noise, informasi yang tidak relevan, dan kesalahan, sehingga data yang digunakan dalam analisis lebih bersih, akurat, dan dapat diandalkan dalam analisis.

Tokenizing: *Tokenizing* adalah tahapan kedua pada proses pendekatan NLP dimana proses ini akan membagi teks menjadi token, seperti kata atau kalimat. Hal ini memudahkan pemrosesan lebih lanjut. Adapun *library* yang digunakan yakni menggunakan NLTK (*Natural Language Toolkit*) [4].

Proses *Stemming* digunakan untuk mengubah kata-kata ke bentuk dasar atau kata kerja infinitif. Misalnya, kata "berlari," "berlari-lari," dan "lari" dapat diubah menjadi "lari" [11]. *Stemming* membantu mengurangi variasi kata yang memiliki arti yang sama. *Library* yang digunakan pada *stemming* adalah sastrawi yang dapat mengubah kata-kata berimbuhan dalam Bahasa Indonesia menjadi bentuk kata dasar.

TF-IDF merupakan proses yang digunakan untuk melakukan transformasi data dari sebuah teks menjadi sebuah numerik guna dilakukan untuk pembobotan pada tiap kata [12]. Teknik TF-IDF (Term Frequency-Inverse Document Frequency) memperkirakan seberapa penting sebuah kata dalam dokumen atau koleksi dokumen. Rumus kalkulasi TF-IDF menggabungkan frekuensi kata dalam dokumen dengan inversi frekuensi kata dalam keseluruhan dokumen, sehingga memberikan bobot yang lebih tinggi pada kata-kata yang jarang muncul tapi memiliki makna penting [3].

2.4. Klasifikasi User Review Menggunakan Algoritma KNN

Algoritma K-Nearest Neighbor atau KNN merupakan algoritma dalam melakukan klasifikasi objek berdasarkan jarak data pembelajaran baru berdasarkan K tetangga yang paling dekat [6]. Tahapannya dalam memprediksi

menggunakan algoritma KNN adalah melakukan data training dan data uji berdasarkan kategori pada sampel atau pada data yang lampau dan sesuai dengan k sampel training di mana tetangga terdekat dengan sampel uji, selanjutnya dimasukkan ke dalam kategori yang memiliki probabilitas terbesar [7]. Klasifikasi dalam algoritma K-Nearest Neighbor dilakukan beberapa tahapan menurut [13], di antaranya:

1. Tentukan jumlah k
2. Hitung jarak antara titik data yang diinginkan dan k tetangga menggunakan jarak *euclidean* yang dapat dihitung dengan persamaan (2)

$$D(p,q) = \sqrt{(p_1 - p_2)^2 + (q_1 - q_2)^2} \quad (2)$$

Di mana p dan q merupakan titik data

3. Pilih dan mengurutkan kelompok-kelompok k tetangga terdekat yang memiliki jarak *euclidean* yang terkecil
4. Hitung jumlah titik data dari setiap kelas k tetangga yang dipilih pada tahapan sebelumnya
5. Posisikan titik data baru ke dalam kategori di mana jumlah tetangga maksimum.

2.5. Akurasi

Untuk mengevaluasi performa model klasifikasi, terdapat *confusion matrix* yang membandingkan hasil klasifikasi aktual dengan prediksi yang diberikan oleh model. Ada empat nilai penting dalam matrix confusion: Nilai Positif Benar (TP), Nilai Negatif Benar (TN), Nilai Negatif Benar (FP), dan Nilai Negatif Benar (FN). Proses ini melibatkan perhitungan nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Nilai *precision* digunakan untuk mengukur seberapa akurat model dalam memprediksi data positif, sedangkan nilai *recall* digunakan untuk menghitung seberapa baik model dalam menemukan data positif, dan *f1-score* digunakan untuk mengukur harmonic mean atau keseimbangan antara nilai *precision* dan *recall* [9]. Evaluasi kinerja ini digunakan sebagai parameter untuk mengukur seberapa akurat implementasi metode yang digunakan.

3. Hasil dan Pembahasan

3.1. Data Set

Teknik *crawling* yang dilakukan pada proses pengumpulan data menghasilkan sentimen sebanyak 77 *record* dengan sample sentimen seperti tampak dalam tabel 1 berikut.

Tabel 1. Data Set Hasil Scrapping

No	Sentimen
1	Sampean keren
2	Kadang bisa pake pinjer, kadang bs pake BC. gatauuuu gimana lagi absen wkwwkwk
3	Kok absen gagal terus . Kenapa ya

4	ga bisa dibuka
5	Blm sat set
...	
76	Mantap
77	Mantap

3.2. Preprocessing

Tahapan preprocessing menggunakan pendekatan NLP (*Natural Language Processing*) atau dikenal dengan *text processing*. Adapun tahap yang dilakukan sebagai berikut: *Cleanning*, tokenizing, *stemming*, stopword, dan terakhir TF-IDF. Tahapan tersebut secara sistematis dilakukan agar memperoleh output data yang optimal.

Cleanning: Pada tahap awal ini, data teks yang sudah ada dibersihkan dengan menghilangkan url, simbol, angka, tanda baca, hastag, spasi, yang berlebihan maupun karakter yang tidak ada di alfabet. Dan juga mengubah kata singkatan menjadi kata aslinya. Tabel 2 berikut disajikan simulasi data teks sebelum dan sesudah *cleanning*.

Tabel 2. Proses *Cleanning*

No	Sentimen sebelum <i>cleanning</i>	Sentimen sesudah <i>cleanning</i>
1	Sampean keren	sampean keren
2	Kadang bisa pake pinjer, kadang bs pake BC. gatauuu gimana lagi absen wkwwk	kadang bisa pake pinjer, kadang bs pake bc. gatauuu gimana lagi absen wkwwk
3	Kok absen gagal terus . Kenapa ya	kok absen gagal terus . kenapa ya
4	ga bisa dibuka	ga bisa dibuka
5	Blm sat set	blm sat set
...		
76	Mantap	mantap
77	Mantap	mantap

Tokenizing: Tokenizing merupakan proses pemecahan sekumpulan karakter dalam suatu teks kedalam satuan kata, pada proses ini juga dapat menghilangkan karakter pembatas, menghapus angka, dan menghapus tanda baca atau karakter yang bukan merupakan huruf.

Tabel 3. Proses *Tokenizing*

No	Sentimen sebelum <i>Tokenizing</i>	Sentimen sesudah <i>Tokenizing</i>
1	sampean keren	['sampean', 'keren']
2	kadang bisa pake pinjer, kadang bs pake bc. gatauuu gimana lagi absen wkwwk	['kadang', 'bisa', 'pake', 'pinjer', 'kadang', 'bs', 'pake', 'bc', 'gatauuu', 'gimana', 'lagi', 'absen', 'wkwwk']
3	kok absen gagal terus . kenapa ya	['kok', 'absen', 'gagal', 'terus', 'kenapa', 'ya']
4	ga bisa dibuka	['ga', 'bisa', 'dibuka']
5	blm sat set	['blm', 'sat', 'set']
...		
76	mantap	['mantap']
77	mantap	['mantap']

Stemming, selanjutnya tahapan stemming yang bertujuan untuk membentuk teks menjadi kata dasar atau dikenal dengan istilah term. Proses ini dilakukan dengan menghapus imbuhan berupa awalan dan akhiran. Selanjutnya term tersebut akan didaftar dan diberikan

bobot. Adapun output tahapan stemming dilihat pada Tabel 4.

Tabel 4. Proses *Stemming*

No	Sentimen sebelum <i>Stemming</i>	Sentimen sesudah <i>Stemming</i>
1	sampean keren	['sampean', 'keren']
2	kadang bisa pake pinjer, kadang bs pake bc. gatauuu gimana lagi absen wkwwk	['kadang', 'bisa', 'pake', 'pinjer', 'kadang', 'bs', 'pake', 'bc', 'gatauuu', 'gimana', 'lagi', 'absen', 'wkwwk']
3	kok absen gagal terus . kenapa ya	['kok', 'absen', 'gagal', 'terus', 'kenapa', 'ya']
4	ga bisa dibuka	['ga', 'bisa', 'dibuka']
5	blm sat set	['blm', 'sat', 'set']
...		
76	mantap	['mantap']
77	mantap	['mantap']

Stopword: Proses selanjutnya adalah Stopword yang berujuan untuk penghapusan teks pada list stopwords. Teks yang dihapus merupakan teks yang tidak mewakili isi dokumen, seperti: di, dari, dan, ke, yang dan seterusnya. Tabel 5 berikut menunjukkan simulasi dari tahapan stopwords yang dilakukan.

Tabel 5. Proses *Stopword*

No	Sentimen sebelum <i>Stopword</i>	Sentimen sesudah <i>Stopword</i>
1	sampean keren	['sampean', 'keren']
2	kadang bisa pake pinjer, kadang bs pake bc. gatauuu gimana lagi absen wkwwk	['kadang', 'bisa', 'pake', 'pinjer', 'kadang', 'bs', 'pake', 'bc', 'gatauuu', 'gimana', 'lagi', 'absen', 'wkwwk']
3	kok absen gagal terus . kenapa ya	['kok', 'absen', 'gagal', 'terus', 'kenapa', 'ya']
4	ga bisa dibuka	['ga', 'bisa', 'dibuka']
5	blm sat set	['blm', 'sat', 'set']
...		
76	mantap	['mantap']
77	mantap	['mantap']

TF-IDF: Pada tahap ini dikarenakan data yang dilabelkan menghasilkan imbalance data maka perlu dilakukan handling data, namun dikarenakan data berupa tipe data untaian atau string perlu dilakukan ekstraksi data menjadi tipe data numerik. TF-IDF menjadi metode pilihan untuk melakukan ekstraksi data menjadi tipe data numerik dikarenakan TF-IDF ini mengubah teks menjadi vektor numerik dengan menghitung frekuensi kemunculan kata dalam dokumen dan mengurangi bobot kata yang sering muncul dalam seluruh dokumen.

3.3. Klasifikasi *User Review* Menggunakan Algoritma KNN

Data yang sudah diolah pada tahapan sebelumnya, selanjutnya diimplementasikan pada usulan algoritma *machine learning* yaitu K-Nearest Neighbor atau KNN untuk melakukan klasifikasi *user review* pada Aplikasi SAMPEAN Kota Cirebon dengan *data testing* yang digunakan sebanyak 16 *data testing*. Berikut hasil klasifikasi pengujian pada algoritma KNN dapat dilihat pada Tabel 7.

Tabel 7. Hasil Klasifikasi

Data Testing	Klasifikasi User Reviews
setelah updat malah suka gak bisa ke pengaturan profil tulisannya terjadi keasalahan	Ulasan teknis
mantuuul	Ulasan teknis
ini gimana absen ada tulisan mempros lama banget dan kalo udah lama conect timeout sere terlambat absen nya teru	Ulasan pengalaman pengguna
ok	Ulasan teknis
terimakasih ata responnya saya pernah uninstal aplikasi untuk mendapatkan updat aplikasi namun beberapa kali saya lakukan instal aplikasi malah tidak berhasil masuk mohon arahan dan petunjuk terimakasih kadang bisa pake pinjer kadang bs pake bc gatauuu gimana lagi absen wkwkwk	Ulasan pengalaman pengguna
ini gimana ya kok gk masuk sedangkan nanya k sekolah lain bisa masuk untuk absen gimana penjelasannya	Ulasan pengalaman pengguna
saya masih belum bisa absen barcod load teru	Ulasan pengalaman pengguna
bagu mantap aplikasinya	Ulasan teknis
gagal mane gagal mane	Ulasan teknis
aplikasi sampean di hp saya tidak bisa upload foto jadi kaya rata saja kalau di klik tidak ada perubahan apapun sekarang tidak bisa tersimpan lkh dengan penjelasan minim dokumen sekian sedangkan tautan sama sekali tidak bisa di klik	Ulasan teknis
aplikasi gampang boten ribet bisa bantu keluarga	Ulasan teknis
aplikasi yang sangat bagu mantap sangat mudah digunakan dan user friendli	Ulasan teknis
min barusan saya updat kok jadinya malah ga bisa scan barcod ya blank aja warna item teru aplikasi masih suka ngebug sy instal di hp satunya susah masuk cuma muncul background biru aja sama lambang sampean masukin usernam dan password untuk login udah uninstal instal ulang beberapa kali angger bae kudu priben min	Ulasan pengalaman pengguna
akun saya mi yo kawan mari kita bersama main game ini	Ulasan teknis
min kl mau nanya kata sandi sy lupa kata sandinya	Ulasan teknis

Berdasarkan hasil klasifikasi *user reviews* menggunakan algoritma *machine learning* KNN pada aplikasi SAMPEAN Kota Cirebon dari 16 data testing didapatkan hasil sentimen *user reviews* pada klasifikasi ulasan teknis sebanyak 10 data dan klasifikasi ulasan pengalaman pengguna sebanyak 6 data.

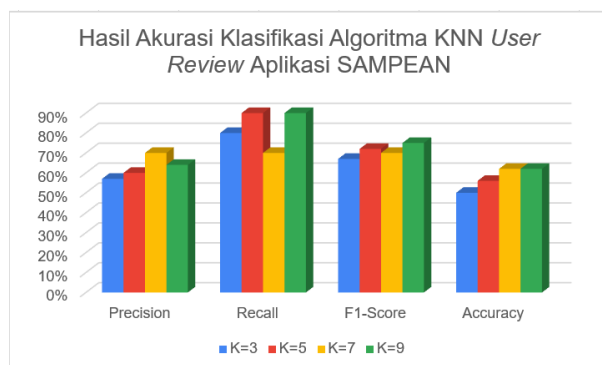
3.6 Akurasi

Akurasi dilakukan menggunakan *confusion matrix* yang diperoleh dari model yang dihasilkan. Berikut hasil nilai *confusion matrix* dari model algoritma KNN.

Tabel 8. Hasil Akurasi

Nilai K	Precision	Recall	F1-Score	Accuracy
K=3	57%	80%	67%	50%
K=5	60%	90%	72%	56%
K=7	70%	70%	70%	62%
K=9	64%	90%	75%	62%

Berdasarkan hasil akurasi menggunakan pengujian *confusion matrix* pada algoritma *machine learning* KNN untuk analisis sentimen *user review* pada aplikasi SAMPEAN Kota Cirebon dengan 4 nilai k, maka dapat disimpulkan (1) Pengujian *Precision* nilai tertinggi yaitu pada k=7, (2) Pengujian *Recall* nilai tertinggi yaitu k=5 dan k=9, (3) Pengujian *F1-Score* nilai tertinggi yaitu k=9, dan (4) Pengujian *Accuracy* nilai tertinggi pada k=7 dan k=9. Berikut Gambar 2. merupakan Visualisai hasil akurasi klasifikasi Algoritma KNN *user review* aplikasi SAMPEAN kota Cirebon.



Gambar 2. Visualisasi Hasil Akurasi

Berdasarkan Gambar 2. Visualisasi akurasi algoritma KNN menggunakan *confusion matrix* dengan pengujian 4 nilai K yaitu K=3, K=5, K=7, dan K=9. terlihat dari visualisasi grafik ini melibatkan perhitungan nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

4. Kesimpulan

Berdasarkan hasil klasifikasi ulasan pengguna menggunakan algoritma *machine learning* KNN pada aplikasi SAMPEAN Kota Cirebon, dari 16 data uji, diperoleh hasil bahwa terdapat 10 data yang diklasifikasikan sebagai ulasan teknis dan 6 data yang diklasifikasikan sebagai ulasan pengalaman pengguna. Hasil ini menunjukkan bahwa klasifikasi sentimen pengguna terhadap aplikasi SAMPEAN Kota Cirebon lebih banyak pada ulasan teknis dibandingkan dengan ulasan pengalaman pengguna. Dari pengujian akurasi menggunakan *confusion matrix* dengan nilai K=3,5,7,9 didapatkan hasil bahwa nilai K=9 memberikan performa terbaik dengan keseimbangan antara *precision*, *recall*, *F1-Score*, dan akurasi. Dengan model mencapai *precision* sebesar 64%, *recall* 90%, *F1-Score* 75%, dan akurasi 62%. Meskipun *precision* sedikit menurun dibandingkan K=7, *recall* dan *F1-Score* tetap tinggi, menunjukkan bahwa model ini mampu mengidentifikasi sebagian besar sampel yang relevan dengan akurasi yang tetap tinggi. Dapat disimpulkan dengan Optimalisasi

parameter K dalam KNN dengan semakin tinggi nilai K maka semakin tinggi tingkat akurasi. Hal ini menegaskan pentingnya pemilihan parameter yang tepat untuk meningkatkan efektivitas model machine learning dalam berbagai aplikasi Natural Language Processing (NLP).

Untuk penelitian di masa depan terkait implementasi optimasi NLP dan KNN untuk sentimen review aplikasi SAMPEAN di Cirebon, beberapa arah pengembangan yang dapat dipertimbangkan meliputi penggunaan teknik NLP yang lebih canggih seperti Transformer-based models (misalnya BERT, GPT) untuk meningkatkan pemahaman konteks dan sentimen dari ulasan pengguna, integrasi dengan deep learning seperti CNN atau RNN untuk meningkatkan akurasi klasifikasi sentimen, analisis multibahasa untuk mendukung berbagai latar belakang bahasa pengguna, penggunaan data eksternal seperti media sosial untuk memperkaya analisis sentimen, optimasi hyperparameter yang lebih mendalam, evaluasi dengan dataset yang lebih besar dan beragam untuk memastikan generalisasi model, pengembangan antarmuka pengguna yang memungkinkan monitoring sentimen secara real-time, serta analisis sentimen berdasarkan konteks untuk melihat perubahan sentimen dalam berbagai situasi.

Daftar Rujukan

- [1] K. T. Hermawan, I. G. Pusparani, and D. Solihudin, "Transformasi Digital Layanan Kepegawaian Pemerintah Daerah Kota Cirebon: Studi Kasus Kebijakan Sistem Administrasi Manajemen Pemerintahan (SAMPEAN)," *J. Stud. Kebijakan. Publik*, vol. 2, no. 1, pp. 13–26, 2023, doi: 10.21787/jskp.2.2023.13-26.
- [2] E. Sera, H. Hazriani, M. Mirfan, and Y. Yuyun, "Analisis Sentimen Ulasan Produk di E-Commerce Bukalapak Menggunakan Natural Language Processing," *Pros. SISFOTEK*, pp. 237–243, 2023, [Online]. Available: <http://www.seminar.iaii.or.id/index.php/SISFOTEK/article/view/406%0Ahttp://www.seminar.iaii.or.id/index.php/SISFOTEK/article/download/406/338>
- [3] N. Azriansyah, E. Indra, and N. Azriansyah, "Penerapan Natural Language Processing Untuk Analisis Sentimen Terhadap Aplikasi Streaming," *J. Ilm. BETRIK (Besemah Teknol. Inf. dan Komputer)*, vol. 14, no. 2, pp. 273–282, 2023, [Online]. Available: <https://ejournal.pppmitpa.or.id/index.php/betrik/article/view/96>
- [4] Muhammad Daffa Al Fahreza, Ardytha Luthfiarta, Muhammad Rafid, and Michael Indrawan, "Analisis Sentimen: Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 16–25, 2024, doi: 10.52158/jacost.v5i1.715.
- [5] A. N. S. Rahayu, T. I. Hermanto, and I. M. Nugroho, "Sentiment Analysis Using K-Nearest Neighbor Based on Particle Swarm Optimization According To Sunscreen'S Reviews," *J. Tek. Inform.*, vol. 3, no. 6, pp. 1639–1646, 2022, doi: 10.20884/1.jutif.2022.3.6.425.
- [6] S. P. Dewi, N. Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- [7] Al-Khowarizmi, R. Syah, M. K. M. Nasution, and M. Elveny, "Sensitivity of MAPE using detection rate for big data forecasting crude palm oil on k-nearest neighbor," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 3, pp. 2696–2703, 2021, doi: 10.11591/ijece.v11i3.pp2696-2703.
- [8] P. Arsi, I. Prayoga, and M. H. Asyari, "Klasifikasi Sentimen Publik Terhadap Jenis Vaksin Covid-19 yang Tersertifikasi WHO Berbasis NLP dan KNN," *J. Media Inform. Budidarma*, vol. 7, no. 1, pp. 260–266, 2023, doi: 10.30865/mib.v7i1.5418.
- [9] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 302–307, 2023, doi: 10.30591/jpit.v8i3.5734.
- [10] S. Rahayu, Y. MZ, J. E. Bororing, and R. Hadiyat, "Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP," *Edumatic J. Pendidik. Inform.*, vol. 6, no. 1, pp. 98–106, 2022, doi: 10.29408/edumatic.v6i1.5433.
- [11] D. Mustikasari, I. Widaningrum, R. Arifin, and W. H. E. Putri, "Comparison of Effectiveness of Stemming Algorithms in Indonesian Documents," *Proc. 2nd Borobudur Int. Symp. Sci. Technol. (BIS-STE 2020)*, vol. 203, pp. 154–158, 2021, doi: 10.2991/aer.k.210810.025.
- [12] M. Prasetya, M. Wulandari, and S. A. Nikmah, "Implementasi NLP (Natural Language Processing) Dasar pada Analisis Sentiment Review Spotify," *Stain. (Seminar Nas. Teknol. Sains)*, vol. 3, no. 1, pp. 145–153, 2024.
- [13] X. Li, J. Zhang, and F. Safara, "Improving the Accuracy of Diabetes Diagnosis Applications through a Hybrid Feature Selection Algorithm," *Neural Process. Lett.*, vol. 55, no. 1, pp. 153–169, 2023, doi: 10.1007/s11063-021-10491-0.