

ANALISIS SENTIMEN ULASAN PENGGUNA ALLO BANK MENGGUNAKAN K-NEAREST NEIGHBORS DAN MENENTUKAN NILAI K OPTIMAL

Marlina¹, Antonius Indra Dharma Prasetya²

¹Program Studi Sistem Informasi, STMIK Kharisma Makassar

¹Jl. Baji Ateka, Makassar – Sulawesi Selatan, Indonesia

²PSDKU Informatika, Universitas Ciputra Surabaya – Indonesia

²CitraLand CBD Boulevard, Surabaya – Jawa Timur, Indonesia

Email: ¹marlina@kharisma.ac.id, ²anindhpra@gmail.com

ABSTRAK

Analisis sentimen merupakan teknik penting dalam memahami opini publik terhadap produk atau layanan, termasuk dalam konteks perbankan digital seperti Allo Bank. Sebagai bank digital yang baru beroperasi, Allo Bank perlu memahami persepsi dan kebutuhan pelanggan untuk membangun kepercayaan serta meningkatkan kualitas layanan. Penelitian ini bertujuan untuk mengklasifikasikan ulasan pengguna Allo Bank yang terdapat di Play Store ke dalam kategori sentimen positif, negatif, atau netral, serta untuk menentukan nilai K optimal pada algoritma K-Nearest Neighbors (KNN) melalui berbagai teknik optimasi. Selain itu, kinerja model KNN dievaluasi menggunakan metrik akurasi, presisi, dan *recall*. Hasil penelitian menunjukkan bahwa 56,4% ulasan bersifat positif, mencerminkan tingkat kepuasan pengguna, sementara 38,8% ulasan negatif mengindikasikan area yang perlu perbaikan, dan hanya 4,8% ulasan yang bersifat netral. Penelitian ini juga menemukan bahwa nilai K optimal dalam algoritma KNN bervariasi tergantung pada teknik optimasi yang digunakan, yaitu *cross-validation*, *grid search*, *random search*, dan *elbow method*. Variasi ini terjadi karena setiap teknik memiliki tujuan, metrik evaluasi, dan ruang pencarian hyperparameter yang berbeda. Oleh karena itu, perlu dilakukan analisis lebih lanjut untuk menentukan nilai K yang paling sesuai dengan karakteristik data dan tujuan penelitian. Selain itu, hasil analisis menunjukkan bahwa kinerja teknik optimasi sangat dipengaruhi oleh proporsi data latih dan data uji. Teknik *cross-validation* memberikan hasil terbaik pada proporsi data uji: data latih sebesar 80:20, berdasarkan metrik evaluasi seperti akurasi, presisi, dan *recall*. Namun, pada proporsi 70:30, tidak ada teknik optimasi yang secara konsisten lebih unggul. Temuan ini menunjukkan bahwa pemilihan teknik optimasi perlu disesuaikan dengan konteks dan karakteristik data untuk memastikan performa model yang optimal.

Kata kunci: Allo Bank, Analisa sentimen, K-Nearest Neighbors, KNN, K-Optimal

ABSTRACT

Sentiment analysis is an important technique in understanding public opinion towards products or services, including in the context of digital banking such as Allo Bank. As a newly operating digital bank, Allo Bank needs to understand customer perceptions and needs to build trust and improve service quality. This study aims to classify Allo Bank user reviews on the Play Store into positive, negative, or neutral sentiment categories, and to determine the optimal K value in the K-Nearest Neighbors (KNN) algorithm through various optimization techniques. In addition, the performance of the KNN model was evaluated using accuracy, precision, and recall metrics. The results showed that 56.4% of reviews were positive, reflecting the level of user satisfaction, while 38.8% of negative reviews indicated areas that needed improvement, and only 4.8% of reviews were neutral. This study also found that the optimal K value in the KNN algorithm varies depending on the optimization technique used, namely cross-validation, grid search, random search, and elbow method. This variation occurs because each technique has different objectives, evaluation metrics, and hyperparameter search spaces. Therefore, further analysis is needed to determine the K value that best suits the characteristics of the data and the objectives of the study. In addition, the analysis results show that the performance of optimization techniques is greatly influenced by the proportion of training data and test data. The cross-validation technique gives the best results at a proportion of test data:training data of 80:20, based on evaluation metrics such as accuracy, precision, and recall. However, at a proportion of 70:30, no optimization technique is consistently superior. This finding suggests that the selection of optimization techniques needs to be adjusted to the context and characteristics of the data to ensure optimal model performance.

Keywords: Allo Bank, Sentiment analysis, K-Nearest Neighbors, KNN, K-Optimal

1. PENDAHULUAN

Analisis sentimen merupakan metode yang digunakan untuk mengevaluasi opini publik terhadap produk atau layanan tertentu, termasuk aplikasi keuangan seperti Allo Bank. Dalam konteks ini, algoritma K-Nearest Neighbors (KNN) menjadi salah satu teknik yang efektif untuk mengklasifikasikan ulasan pengguna berdasarkan sentimen positif, negatif, atau netral. K-NN bekerja dengan cara mengidentifikasi kedekatan data baru dengan data yang sudah ada, sehingga memungkinkan untuk mengklasifikasikan sentimen berdasarkan karakteristik yang ada dalam data pelatihan [1]. Dalam penelitian yang dilakukan oleh Putra dan Juanita, analisis sentimen menggunakan KNN menunjukkan bahwa metode ini mampu mengekstrak opini publik dari teks yang tidak terstruktur, seperti ulasan aplikasi [2]. Selain itu, Puspitasari juga menunjukkan bahwa KNN dapat digunakan untuk menganalisis sentimen dalam konteks yang lebih luas, seperti inflasi pasca-COVID-19, dengan hasil akurasi yang cukup baik [3]. Hal ini menunjukkan bahwa KNN tidak hanya relevan untuk aplikasi spesifik tetapi juga dapat diadaptasi untuk berbagai konteks analisis sentimen. Lebih lanjut, Maharani dan Triayudi juga menekankan pentingnya pemilihan fitur dalam analisis sentimen, di mana mereka menggunakan KNN untuk mengevaluasi kepuasan pelanggan terhadap layanan pembayaran digital di Indonesia [4]. Penelitian ini menunjukkan bahwa metode KNN dapat digunakan untuk menganalisis sentimen dalam konteks layanan keuangan, yang sangat relevan dengan tujuan analisis sentimen terhadap Allo Bank. Dengan memanfaatkan dataset yang diambil dari platform media sosial, analisis ini dapat memberikan wawasan yang berharga tentang persepsi pengguna terhadap aplikasi tersebut. Dari beberapa teknik analisis data dalam data science, analisis sentimen digunakan karena dapat menghemat waktu analisis data untuk mengetahui tingkat sentimen suatu topik [5].

Dalam konteks perbankan digital, seperti Allo Bank, analisis sentimen dapat memberikan wawasan berharga mengenai kepuasan pelanggan dan area yang perlu diperbaiki. Allo Bank, sebagai salah satu pelopor dalam layanan perbankan digital di Indonesia, menghadapi tantangan untuk memahami dan memenuhi harapan pengguna. Ulasan pengguna di platform digital sering kali mencerminkan pengalaman dan kepuasan mereka terhadap layanan yang diberikan. Oleh karena itu, analisis sentimen terhadap ulasan pengguna dapat membantu manajemen dalam mengambil keputusan strategis untuk meningkatkan kualitas layanan. Analisis sentimen menggunakan KNN dapat memberikan gambaran yang jelas tentang bagaimana pengguna merasakan layanan Allo Bank. Dengan memanfaatkan teknik-teknik pemrosesan data yang tepat dan algoritma KNN, penelitian ini dapat membantu dalam pengambilan keputusan strategis untuk meningkatkan layanan dan kepuasan pengguna di masa mendatang [6]. Dengan meningkatnya penggunaan aplikasi perbankan di Indonesia, penting untuk menganalisis bagaimana pengguna merespons layanan yang diberikan. Metode KNN adalah salah satu teknik yang efektif untuk melakukan klasifikasi sentimen, karena kemampuannya dalam menangani data yang tidak terstruktur dan memberikan hasil yang akurat dalam pengelompokan data berdasarkan kedekatan fitur [7][8] serta telah terbukti efektif dalam berbagai konteks penelitian.

Penelitian ini bertujuan untuk mengklasifikasikan ulasan pengguna Allo Bank ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Selain itu, penelitian ini bertujuan menentukan nilai K optimal dalam algoritma KNN menggunakan empat teknik optimasi, yaitu Cross-Validation, Grid Search, Random Search, dan Elbow Method. Penelitian ini juga menganalisis pengaruh perbedaan pembagian dataset (80:20 dan 70:30) terhadap hasil klasifikasi, serta mengevaluasi teknik optimasi yang memberikan akurasi terbaik.

Hasil penelitian ini diharapkan dapat memberikan rekomendasi yang lebih baik dalam pengembangan produk dan layanan Allo Bank, sehingga mampu meningkatkan kepuasan dan loyalitas pengguna. Selain itu, penelitian ini diharapkan dapat menjadi referensi bagi penelitian serupa di masa mendatang, khususnya dalam penerapan algoritma KNN untuk analisis sentimen di sektor perbankan. Penentuan nilai k yang optimal dalam algoritma KNN sangat penting untuk meningkatkan akurasi model dalam mengklasifikasikan sentimen pengguna. Dengan memilih k yang tepat, model dapat meminimalkan kesalahan prediksi dan menghasilkan analisis sentimen yang lebih andal, memberikan wawasan yang lebih baik untuk pengembangan produk dan layanan Allo Bank.

2. METODE PENELITIAN

Prinsip Kerja KNN

Algoritma K-Nearest Neighbors (KNN) adalah salah satu metode klasifikasi yang sederhana namun sangat efektif dalam analisis data. Algoritma ini bekerja dengan cara mengklasifikasikan data baru berdasarkan kedekatannya dengan data yang sudah ada dalam dataset. KNN mengandalkan prinsip bahwa objek yang memiliki karakteristik serupa cenderung berada dekat satu sama lain dalam ruang fitur. Proses kerja KNN dimulai dengan pemilihan nilai K, yang merupakan jumlah tetangga terdekat yang akan dipertimbangkan dalam klasifikasi. Nilai K yang optimal sangat penting, karena dapat mempengaruhi akurasi model. Jika K terlalu kecil, model dapat menjadi sensitif terhadap noise dalam data, sedangkan jika K terlalu besar, model dapat kehilangan informasi penting tentang kelas minoritas [9]. Setelah nilai K ditentukan, algoritma akan menghitung jarak antara objek yang akan diklasifikasikan dengan semua objek dalam dataset pelatihan. Jarak ini biasanya dihitung menggunakan metrik seperti Euclidean distance, Manhattan distance, atau Minkowski distance [10]. Setelah jarak dihitung, KNN akan memilih K objek terdekat dan melakukan voting untuk menentukan kelas dari objek baru. Kelas yang

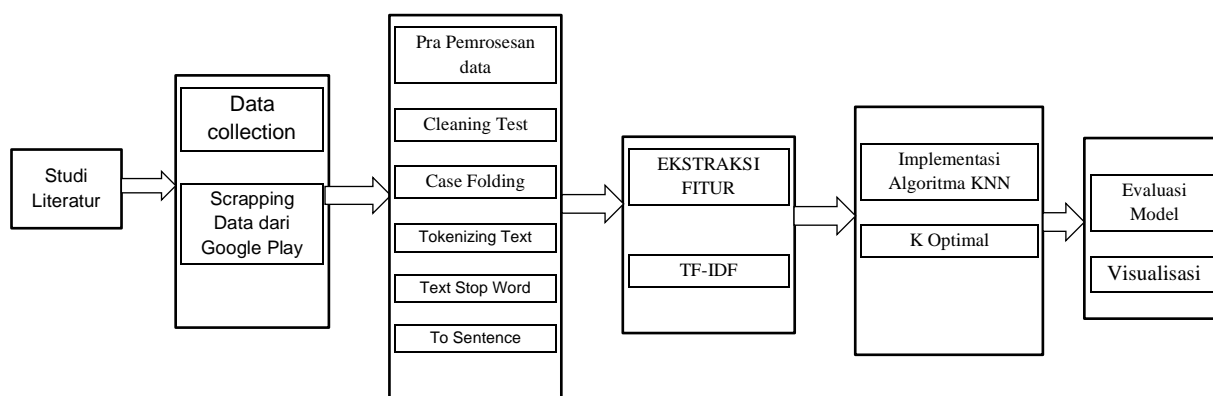
paling banyak dipilih oleh tetangga terdekat akan menjadi kelas dari objek tersebut. Proses ini membuat KNN menjadi metode yang intuitif dan mudah dipahami, serta tidak memerlukan asumsi distribusi data yang kuat.

Ada 4 teknik optimasi digunakan untuk menentukan nilai K optimal yaitu:

1. *Cross-Validation*
Teknik ini membagi dataset menjadi beberapa subset untuk melatih dan menguji model secara bergantian. K-fold cross-validation, misalnya, membagi data menjadi K bagian, di mana setiap bagian digunakan sebagai data uji [11].
2. *Grid Search*
Metode ini melakukan pencarian sistematis melalui kombinasi parameter yang telah ditentukan, termasuk nilai K. Grid search menguji setiap kombinasi untuk menemukan kombinasi terbaik berdasarkan metrik evaluasi yang dipilih [12].
3. *Random Search*
Berbeda dengan grid search, random search memilih kombinasi parameter secara acak dari ruang pencarian yang telah ditentukan. Metode ini seringkali lebih efisien dalam menemukan parameter optimal dalam waktu yang lebih singkat [13].
4. *Elbow Method*
Metode ini digunakan untuk menentukan jumlah cluster yang optimal dalam analisis cluster, tetapi juga dapat diterapkan untuk KNN. Dengan menghitung error rate untuk berbagai nilai K, titik "siku" pada grafik error rate versus nilai K menunjukkan nilai K yang optimal [14].

Metodologi Penelitian

Metodologi penelitian ini dirancang untuk memberikan gambaran sistematis mengenai langkah-langkah yang diambil dalam proses penelitian, mulai dari pengumpulan data hingga analisis hasil.



Gambar 1. Metodologi Penelitian

Pada Gambar 1 merupakan metodologi yang digunakan dalam penelitian ini disusun dalam beberapa langkah utama yaitu:

1. Proses diawali dengan studi literatur untuk memahami konsep dasar, algoritma KNN, teknik optimasi *hyperparameter*, serta pendekatan dalam analisis sentimen.
2. Selanjutnya, dilakukan pengumpulan dataset ulasan pengguna Allo Bank melalui play store dengan melakukan Scrapping data menggunakan google-play-scraper pandas.
3. Pra-pemrosesan data dengan melakukan:
 - 1) Pembersihan Data (*Cleaning Text*): teks asli telah dibersihkan dengan menghapus mention, hashtag, tautan, angka, dan karakter non-huruf, sehingga hanya menyisakan kata-kata yang relevan.
 - 2) Case Folding: Seluruh huruf kapital diubah menjadi huruf kecil untuk konsistensi dalam pemrosesan.
 - 3) Tokenisasi Teks (*Tokenizing Text*): teks telah dipecah menjadi daftar token, yaitu kata-kata individu yang menyusun kalimat.
 - 4) Penyaringan Teks (*Filtering Text*): stopwords, atau kata-kata umum yang tidak memiliki banyak makna, seperti iya, yaa, gak, dan, telah dihapus dari teks.
 - 5) Stemming Teks (*Stemming Text*): kata-kata dengan imbuhan telah diubah ke bentuk akar untuk menyederhanakan analisis.
 - 6) Konversi ke Kalimat (*To Sentence*): token-token yang tersisa diolah kembali menjadi kalimat yang terstruktur.

4. Ekstraksi Fitur

Ekstraksi fitur TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah salah satu metode yang sering digunakan dalam Natural Language Processing (NLP) untuk merepresentasikan teks dalam bentuk numerik. Representasi ini digunakan untuk mengetahui seberapa penting suatu kata dalam sebuah dokumen relatif terhadap kumpulan dokumen lainnya. Untuk menghitung TF dan IDF [15].

$$TF(t,d) = \frac{f(t,d)}{N}$$

Keterangan:

- Term Frequency (TF): Mengukur seberapa sering kata t muncul dalam dokumen d .
- $f(t, d)$ adalah jumlah kemunculan kata t dalam dokumen d .
- N adalah jumlah total kata dalam dokumen d .

$$IDF(t,D) = \log \left(\frac{Nd}{1+df(t,D)} \right)$$

Keterangan:

- Inverse Document Frequency (IDF): mengukur seberapa unik kata t di seluruh kumpulan dokumen D .
- Nd adalah jumlah dokumen dalam kumpulan dokumen D .
- $df(t,D)$ adalah jumlah dokumen yang mengandung kata t .

5. Implementasi Algoritma KNN

Implementasikan algoritma KNN dengan nilai K Optimal menggunakan teknik optimasi Cross-Validation, Grid Search, Random dan Elbow Method dalam 2 skenario yaitu memisahkan dataset menjadi 80:20, dan 70:30.

6. Evaluasi Model dan Visualisasi

Melakukan evaluasi performa model untuk setiap nilai k selanjutnya dilakukan visualisasi kedalam grafik. Adapun evaluasi model menggunakan:

- 1) Akurasi : mengukur proporsi prediksi yang benar (baik positif maupun negatif) dari total data yang diuji.
- 2) Presisi : proporsi data yang diprediksi sebagai positif (atau kategori tertentu) yang benar-benar positif.
- 3) Recall : proporsi data yang benar-benar positif yang berhasil diprediksi dengan benar oleh model.

3. HASIL DAN PEMBAHASAN

Pada bagian ini, akan disajikan hasil dari analisis sentimen yang dilakukan menggunakan algoritma K-Nearest Neighbors (KNN) dengan berbagai nilai K optimal. Pembahasan akan difokuskan pada interpretasi hasil yang diperoleh serta perbandingan performa model berdasarkan metrik evaluasi yaitu akurasi, presisi, *recall*. Adapun hasil penelitian sebagai berikut:

1. Data yang digunakan dalam penelitian ini berasal dari ulasan pengguna aplikasi Allo Bank yang diperoleh melalui platform Play Store. Proses pengumpulan data dilakukan menggunakan teknik web scraping, yang memungkinkan ekstraksi informasi relevan dari ulasan pengguna secara efisien.
2. Data yang terkumpul selanjutnya akan melalui tahap pembersihan dan prapemrosesan untuk memastikan kualitasnya sebelum dianalisis. Analisis ini bertujuan untuk mengevaluasi sentimen pengguna dan mengidentifikasi area perbaikan yang dapat meningkatkan pengalaman pengguna. Metode text mining akan diterapkan untuk menggali wawasan yang dapat digunakan sebagai dasar dalam pengambilan keputusan strategis terkait pengembangan aplikasi. Hasil akhir dari proses ini ditampilkan pada Tabel 1, di mana data mentah diubah menjadi format yang lebih terstruktur dan siap untuk analisis lanjutan. Proses ini memungkinkan evaluasi sentimen dilakukan dengan lebih akurat dan efisien.

Tabel 1. Pra pemrosesan Data

content	score	text_clean	text_casefoldingText	text_tokenizingText	text_stopword	text_akhir
Terima kasih allo bank, uang saya sudah kembal...	5	Terima kasih allo bank uang saya sudah kembali...	terima kasih allo bank uang saya sudah kembali...	[terima, kasih, allo, bank, uang, saya, sudah,...	[terima, kasih, allo, bank, uang, allo, bank, ...	terima kasih allo bank uang allo bank terbaik
Tes aplikasinya dulu	3	Tes aplikasinya dulu	tes aplikasinya dulu	[tes, aplikasinya, dulu]	[tes, aplikasinya]	tes aplikasinya

content	score	text_clean	text_casefoldingText	text_tokenizingText	text_stopword	text_akhir
Min kenapa verifikasi data saya lama banget ud...	2	Min kenapa verifikasi data saya lama banget ud...	min kenapa verifikasi data saya lama banget ud...	[min, kenapa, verifikasi, data, saya, lama, ba...	[min, verifikasi, data, banget, udh, nunggu, h...	min verifikasi data banget udh nunggu hubungi ...
APLIKASI TIDAK JELAS !!! LUPA PASSWORD TAPI TI...	1	APLIKASI TIDAK JELAS LUPA PASSWORD TAPI TIDAK...	aplikasi tidak jelas lupa password tapi tidak...	[aplikasi, tidak, jelas, lupa, password, tapi,...	[aplikasi, lupa, password, menerima, kode, otp...	aplikasi lupa password menerima kode otp nomor...
Apk ga jelas , Waktu mau VerMuk GELAP BANGET K...	1	Apk ga jelas Waktu mau VerMuk GELAP BANGET KA...	apk ga jelas waktu mau vermuk gelap banget ka...	[apk, ga, jelas, waktu, mau, vermuk, gelap, ba...	[apk, vermuk, gelap, banget, kaya, pake, kamer...	apk vermuk gelap banget kaya pake kamera rusak...
salah satu aplikasi paylater yang simple dan ...	5	salah satu aplikasi paylater yang simple dan ...	salah satu aplikasi paylater yang simple dan ...	[salah, satu, aplikasi, paylater, yang, simpl...	[salah, aplikasi, paylater, simple]	salah aplikasi paylater simple

- Untuk mendukung proses analisis, dilakukan ekstraksi fitur dengan mengubah teks menjadi representasi numerik menggunakan teknik TF-IDF (*Term Frequency-Inverse Document Frequency*). Teknik ini memungkinkan identifikasi kata-kata penting dalam sebuah dokumen dengan mempertimbangkan frekuensi kemunculannya di dokumen tertentu dibandingkan dengan seluruh dokumen dalam dataset. Representasi numerik ini sangat penting untuk memastikan algoritma dapat memahami dan memproses data teks secara efektif. Dalam penelitian ini, penulis mengaplikasikan teknik TF-IDF menggunakan implementasi standar yang tersedia pada library Scikit-learn (Sklearn).
- Visualisasi wordcloud yang dihasilkan dapat dilihat Gambar 2 yaitu menampilkan kata-kata yang paling sering muncul dalam ulasan, dengan ukuran yang mencerminkan frekuensi kemunculannya. Wordcloud ini membantu mengidentifikasi kata-kata kunci yang berkontribusi pada masing-masing kategori sentimen. Untuk melakukan klasifikasi sentimen pada penilaian, penulis menetapkan kategori berdasarkan nilai rating yang diberikan. Penilaian dengan rating lebih dari 3 dikategorikan sebagai positif, mencerminkan respon yang baik atau puas. Penilaian dengan rating tepat 3 dikategorikan sebagai netral, menunjukkan respon yang tidak condong pada sisi positif maupun negatif. Sementara itu, penilaian dengan rating kurang dari 3 dikategorikan sebagai negatif, mencerminkan ketidakpuasan atau respon kurang baik. Klasifikasi ini digunakan untuk mendukung analisis sentimen secara lebih terarah dan bermakna. Hasil yang diperoleh adalah Dari hasil pelabelan diperoleh ulasan Positif sebesar 56,4%, ulasan Negatif sebesar 38,8% dan ulasan netral sebesar 4,8%.



Gambar 2. Pelabelan Data

5. Implementasi metode KNN:

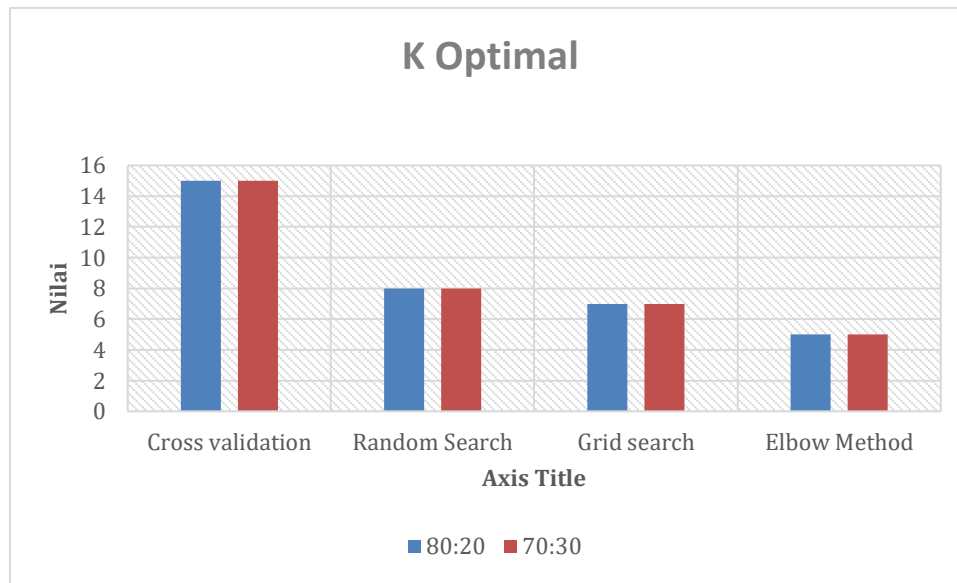
- Dalam penelitian ini, dilakukan pembagian data uji dan data latih untuk memastikan model memiliki performa yang optimal dalam melakukan klasifikasi. Penulis menerapkan dua skenario pembagian data, yaitu 80:20 dan 70:30, di mana masing-masing skenario menunjukkan persentase data latih

dan data uji. Pembagian ini bertujuan untuk menguji konsistensi dan keakuratan model pada berbagai proporsi data, sehingga hasil evaluasi dapat lebih komprehensif.

2) Gambar 3 menghasilkan nilai K optimal menggunakan teknik:

- a. Cross-Validation menghasilkan K optimal=15
- b. Grid Search menghasilkan K optimal = 8
- c. Random Search menghasilkan K optimal=7
- d. Elbow Method menghasilkan K optimal=5

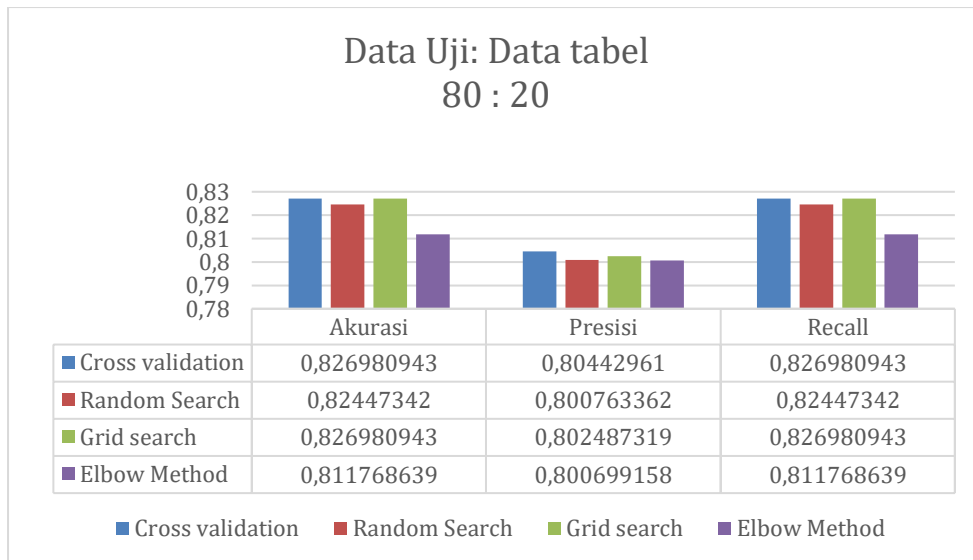
Hasil penelitian menunjukkan bahwa variasi dataset tidak memberikan pengaruh signifikan terhadap kinerja model KNN, namun nilai K optimal yang diperoleh dari setiap teknik optimasi (*cross-validation*, *grid search*, *random search*, dan *elbow method*) berbeda-beda.



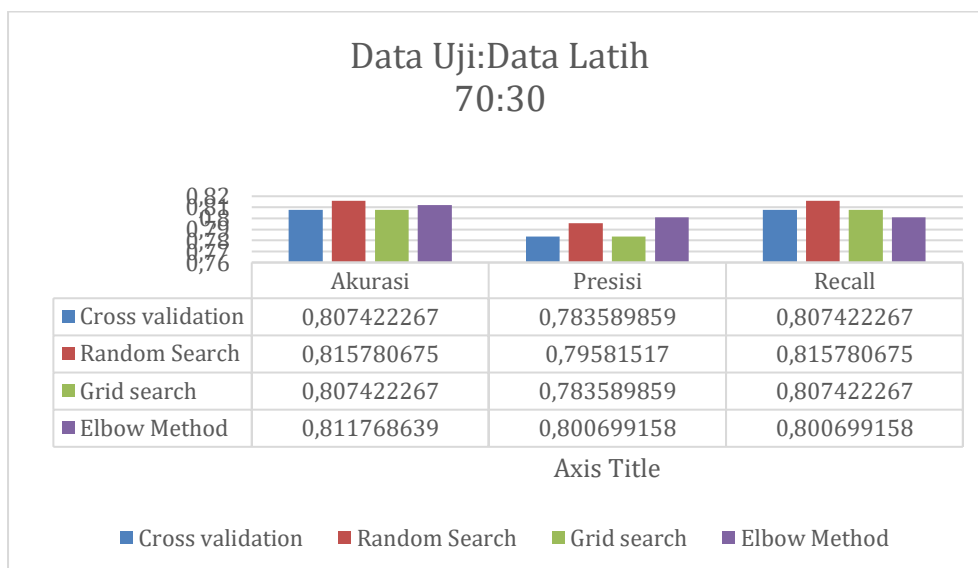
Gambar 3. Penentuan K optimal

6. Evaluasi Model dan Visualisasi

Gambar 4 menunjukkan bahwa *cross-validation* secara konsisten menghasilkan estimasi performa model yang lebih akurat dibanding teknik lainnya karena memanfaatkan seluruh dataset secara bergantian untuk pelatihan dan pengujian. *Random search*, meskipun efisien dalam eksplorasi hyperparameter, cenderung melewati kombinasi optimal, terutama pada dataset dengan ukuran besar atau kompleksitas tinggi. *Grid search*, meskipun menyeluruh, memiliki kelemahan berupa waktu komputasi yang tinggi dan risiko *overfitting* ketika jumlah hyperparameter yang diuji terlalu banyak. Sebaliknya, *elbow method* lebih relevan untuk masalah *unsupervised learning*, seperti menentukan jumlah cluster dalam *K-Means*, sehingga kurang cocok untuk evaluasi berbasis akurasi, presisi, dan *recall*. Oleh karena itu, keunggulan *cross-validation* dalam menjaga keseimbangan antara bias dan variansi membuatnya menjadi pilihan terbaik dalam pengujian model *supervised learning* dengan proporsi data latih 80:20.



Gambar 4. Nilai Akurasi, Presisi dan *Recall* untuk dataset 80:20



Gambar 5. Nilai Akurasi, Presisi dan *Recall* untuk dataset 70:30

Gambar 5 memperlihatkan pada data latih 70:30, random search menawarkan fleksibilitas dalam eksplorasi hyperparameter, tetapi hasilnya tidak selalu konsisten karena bergantung pada kombinasi parameter yang diuji. Dalam model KNN, pemilihan nilai K yang tepat sangat penting untuk menghindari sensitivitas terhadap noise atau underfitting. Meskipun variasi dataset tidak signifikan, pengaturan nilai K yang optimal tetap menjadi faktor kunci dalam mencapai kinerja terbaik.

4. SIMPULAN

Hasil penelitian menunjukkan bahwa persentase ulasan positif yang mencapai 56,4% mencerminkan bahwa sebagian besar pengguna merasa puas dengan layanan yang diberikan oleh Allo Bank. Namun, terdapat 38,8% ulasan negatif, yang mengindikasikan masih adanya aspek layanan yang perlu ditingkatkan untuk memenuhi ekspektasi pengguna secara keseluruhan. Sementara itu, ulasan netral yang hanya sebesar 4,8% menunjukkan bahwa sebagian besar pengguna memiliki pendapat yang tegas, baik dalam sisi positif maupun negatif terhadap layanan Allo Bank.

Selain itu, analisis terhadap nilai K optimal yang dihasilkan dari berbagai teknik optimasi, seperti cross-validation, grid search, random search, dan elbow method, menunjukkan variasi yang signifikan. Perbedaan ini disebabkan oleh faktor-faktor seperti tujuan teknik, metrik evaluasi yang digunakan, serta ruang pencarian

hyperparameter yang berbeda untuk masing-masing metode. Oleh karena itu, penting untuk melakukan analisis lebih lanjut guna menentukan nilai K yang paling sesuai dengan karakteristik data dan tujuan penelitian.

Dalam analisis komparatif, hasil menunjukkan bahwa kinerja teknik optimasi sangat dipengaruhi oleh proporsi data latih. Pada proporsi 80:20, teknik cross-validation memberikan hasil terbaik berdasarkan metrik evaluasi yang digunakan. Namun, pada proporsi 70:30, tidak ada teknik tunggal yang secara konsisten menunjukkan performa dominan. Hal ini mengindikasikan bahwa pemilihan teknik optimasi sebaiknya mempertimbangkan proporsi data serta tujuan analisis secara menyeluruh.

DAFTAR PUSTAKA

- [1] S. Rahayu, Y. Mz, J. Bororing, and R. Hadiyat, "Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial Flip", *Edumatic: Jurnal Pendidikan Informatika*, vol. 6, no. 1, p. 98-106, 2022. <https://doi.org/10.29408/edumatic.v6i1.5433>.
- [2] A. Putra and S. Juanita, "Analisis Sentimen pada Ulasan Pengguna Aplikasi Bibit dan Bareksa dengan Algoritma KNN", *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 2, p. 636-646, 2021. <https://doi.org/10.35957/jatisi.v8i2.962>.
- [3] Ratih Puspitasari, Y. Findawati, and M. Rosid, "Sentiment Analysis of Post-Covid-19 Inflation Based on Twitter Using the K-Nearest Neighbor and Support Vector Machine Classification Methods", *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 4, p. 669-679, 2023. <https://doi.org/10.52436/1.jutif.2023.4.4.801>.
- [4] A. Maharani and A. Triayudi, "Sentiment Analysis of Indonesian Digital Payment Customer Satisfaction Towards Gopay, Dana, and Shopeepay Using Naïve Bayes And K-Nearest Neighbour Methods", *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, p. 672, 2022. <https://doi.org/10.30865/mib.v6i1.3545>.
- [5] W. Winarto, I. A. Musdar, and H. Hasniati, "Sentiment analysis of 2024 Presidential Candidate Using the Support Vector Machine Algorithm on Twitter," *KHARISMA Tech*, vol. 19, no. 1, pp. 86–98, 2024. <https://doi.org/10.55645/kharismatech.v19i1.464>.
- [6] S. Ernawati and R. Wati, "Sentiment Analysis of Three-Period Polemics Using K-Nearest Neighbor With Tf-Idf Weighting", *Jurnal Riset Informatika*, vol. 4, no. 3, p. 215-220, 2022. <https://doi.org/10.34288/jri.v4i3.377A>. Sari, T. Hermanto, and M. Defriani, "Sentiment Analysis of Tourist Reviews Using K-Nearest Neighbors Algorithm and Support Vector Machine", *Sinkron*, vol. 8, no. 3, p. 1366-1378, 2023. <https://doi.org/10.33395/sinkron.v8i3.12447>.
- [7] A. Sari, T. Hermanto, and M. Defriani, "Sentiment Analysis of Tourist Reviews Using K-Nearest Neighbors Algorithm and Support Vector Machine", *Sinkron*, vol. 8, no. 3, p. 1366-1378, 2023. <https://doi.org/10.33395/sinkron.v8i3.12447>.
- [8] R. Amardita, A. Adiwijaya, and M. Purbolaksono, "Analisis Sentimen Terhadap Ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung Menggunakan Algoritma KNN", *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 1, p. 62, 2022. <https://doi.org/10.30865/jurikom.v9i1.3793>.
- [9] I. Iswanto, T. Tulus, and P. Poltak, "Comparison of Feature Selection to Performance Improvement of K-Nearest Neighbor Algorithm in Data Classification", *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 6, p. 1709-1716, 2022. <https://doi.org/10.20884/1.jutif.2022.3.6.471>.
- [10] P. Rinanda, B. Delvika, S. Nurhidayarnis, N. Abror, and A. Hidayat, "Perbandingan klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes pada Ibu Hamil", *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 2, p. 68-75, 2022. <https://doi.org/10.57152/malcom.v2i2.432>.
- [11] W. Ji, "Research on Life Expectancy Prediction Based on Logistic Regression and KNN", *BCP Business & Management*, vol. 40, p. 174-183, 2023. <https://doi.org/10.54691/bcpbm.v40i.4378>.
- [12] M. Kabir, M. Islam, A. Kabir, A. Haque, and M. Rhaman, "Detection of Depression Severity Using Bengali Social Media Posts on Mental Health: Study Using Natural Language Processing Techniques", *JMIR Formative Research*, vol. 6, no. 9, p. e36118, 2022. <https://doi.org/10.2196/36118>.
- [13] S. Williams, S. Bariselli, C. Palego, R. Holland, and P. Coss, "A comparison of Machine-Learning Assisted Optical and Thermal Camera Systems for Beehive Activity Counting", *SSRN Electronic Journal*, 2021. <https://doi.org/10.2139/ssrn.3990570>.
- [14] M. Özateş, A. Yaman, F. Salami, S. Campos, S. Wolf, and U. Schneider, "Identification and Interpretation of Gait Analysis Features and Foot Conditions by Explainable Ai", *Scientific Reports*, vol. 14, no. 1, 2024. <https://doi.org/10.1038/s41598-024-56656-4>.
- [15] A. Purpura, D. Giorgianni, G. Orrù, G. Melis, and G. Sartori, "Identifying Single-Item Faked Responses in Personality Tests: a New Tf-Idf-Based Method", *Plos One*, vol. 17, no. 8, p. e0272970, 2022. <https://doi.org/10.1371/journal.pone.0272970>.