LINIER
LITERATUR INFORMATIKA & KOMPUTER
e-ISSN: 3063-2218
DOI : 10.33096/linier.v2i3.3147
Literatur Informatika & Komputer
Vol 2, No 3 (2025), pp 358-366

Implementasi Algoritma *K-Nearest Neighbors* Untuk Klasifikasi Data Ulasan Pengguna Aplikasi Sulselbar *Mobile* pada *Google Play Store*

Muh. Fadhil Attariq Hasril[a], Purnawansyah[b], Lutfi Budi Ilmawan[c]
Universitas Muslim Indonesia, Makassar, Indonesia
[a]13020210209@umi.ac.id; [b]purnawansyah@umi.ac.id; [c]lutfibudi.ilmawan@umi.ac.id

Received: 09-08-2025 | Revised: 20-08-2025 | Accepted: 13-09-2025 | Published: 29-09-2025

Abstrak

Perkembangan teknologi perbankan digital mendorong Bank Sulselbar menghadirkan layanan mobile banking melalui aplikasi Sulselbar Mobile. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan pengguna aplikasi tersebut di Google Play Store menggunakan algoritma *K-Nearest Neighbors* (KNN). Sebanyak 1000 data ulasan dikumpulkan dengan teknik *web scraping*, kemudian dilakukan pelabelan manual dan serangkaian proses *preprocessing* seperti *case folding*, normalisasi, *stemming*, *stopword removal*, dan *tokenizing*. Data kemudian dikonversi menjadi bentuk numerik menggunakan metode TF-IDF. Model KNN dikembangkan dengan nilai $k = 3$ dan divalidasi menggunakan *5-fold cross-validation*. Hasil evaluasi menunjukkan model memiliki performa yang baik dan konsisten, dengan rata-rata akurasi 85,6%, presisi 85,4%, *recall* 85,6%, dan F1-score 85,5%. Fold terbaik mencapai akurasi 90% dan F1-score 89,8%. Hasil ini menunjukkan bahwa kombinasi TF-IDF dan algoritma KNN efektif dalam melakukan klasifikasi sentimen berbasis teks. Temuan ini dapat dimanfaatkan untuk mengevaluasi kualitas layanan aplikasi serta mendukung pengambilan keputusan dalam pengembangan layanan digital perbankan.

Kata kunci: Sentimen analisis, KNN, TF-IDF, Klasifikasi, *Mobile Banking*

Pendahuluan

Perkembangan teknologi informasi dan komunikasi yang pesat telah mendorong perubahan signifikan dalam berbagai aspek kehidupan, termasuk dalam sektor perbankan. Bank adalah lembaga perantara keuangan atau biasa disebut *financial intermediary.* Artinya, lembaga bank adalah lembaga yang dalam aktivitasnya berkaitan dengan masalah uang [1]. Bank merupakan salah satu komponen vital dalam perekonomian suatu negara. Selain berperan sebagai lembaga yang mengedepankan kepercayaan, bank juga berfungsi sebagai perantara keuangan, mendukung kelancaran sistem pembayaran, serta memainkan peran penting sebagai sarana pelaksanaan kebijakan pemerintah, khususnya dalam hal kebijakan moneter [2]. Dengan adanya inovasi di sektor keuangan, *mobile banking* (M-Banking) hadir sebagai salah satu terobosan yang mempermudah konsumen dalam mengakses berbagai layanan perbankan. *M-banking* adalah layanan dalam bentuk aplikasi digital yang dirancang khusus oleh suatu bank untuk memudahkan nasabah dalam melakukan berbagai transaksi melalui *smartphone* [3].

Bank Pembangunan Daerah Sulawesi Selatan dan Sulawesi Barat (Sulselbar) merupakan lembaga keuangan milik pemerintah daerah yang berfokus pada penyediaan layanan perbankan di wilayah Provinsi Sulawesi Selatan dan Provinsi Sulawesi Barat. Bank ini memiliki peran penting dalam mendorong pembangunan ekonomi di daerah tersebut, dengan menyediakan berbagai produk dan layanan perbankan yang berorientasi pada kebutuhan masyarakat setempat. Dengan kemajuan teknologi yang terus berkembang, PT. Bank Sulselbar juga menawarkan layanan *mobile banking* yang telah diluncurkan sejak tahun 2018, sehingga para nasabah dapat menikmati berbagai layanan perbankan Bank Sulselbar secara praktis melalui genggaman tangan [4]. Layanan mobile banking dapat diakses melalui menu yang terdapat dalam aplikasi Sulselbar Mobile, yang bisa diunduh oleh setiap nasabah Bank Sulselbar yang memiliki perangkat smartphone [3]. Aplikasi Sulselbar Mobile merupakan salah satu aplikasi yang dikelola oleh Pemerintah Provinsi Sulawesi Selatan dan Barat (Sulselbar) untuk memberikan berbagai layanan publik secara daring. Aplikasi ini menawarkan berbagai fitur, mulai dari informasi publik hingga layanan administrasi yang dapat diakses oleh masyarakat luas. Saat ini di Google Play Store, bank Sulselbar Mobile memiliki jumlah pengunduh lebih dari 100 ribu pengguna dengan jumlah ulasan 5.484 dengan rating 3,9 dari 5. Hal ini dapat dijadikan dasar untuk analisis sentiment terhadap aplikasi Sulselbar Mobile guna mengetahui apakah jumlah unduhan, total ulasan, dan rating aplikasi benar-benar relevan dalam membuktikan bahwa aplikasi Sulselbar *Mobile* telah sesuai dengan kebutuhan konsumen.

Analisis Sentimen merupakan salah satu metode untuk menggali informasi berupa opini atau pandangan seseorang terhadap suatu peristiwa atau isu tertentu. Teknik ini dapat dimanfaatkan untuk mengetahui opini publik terkait isu tertentu, *tingkat* kepuasan terhadap layanan, efektivitas kebijakan, mendeteksi cyber bullying, memprediksi pergerakan harga saham, hingga menganalisis pesaing berdasarkan data berbasis teks [5]. Dalam penelitian ini juga menggunakan metode algoritma *K-Nearest Neighbor* (K-NN) untuk melakukan analisis sentiment. K-Nearest Neighbor (KNN) adalah metode klasifikasi yang memanfaatkan data latih yang paling menyerupai objek yang sedang dianalisis. Metode ini sederhana namun mampu memberikan hasil klasifikasi dengan akurasi yang cukup tinggi [6].

Berdasarkan penelitian [7], menggunakan metode K-NN untuk menganalisis sentimen data ulasan pengguna aplikasi Flip di Google Play Store dan algoritma klasifikasi yang digunakan adalah K-Nearest Neighbor dengan pembobotan TF-IDF. Yang menghasilkan, sebanyak 77,67% data uji berhasil diklasifikasikan dengan tepat ke dalam kategori ulasan positif, dengan nilai precision dan recall yang tinggi, masing-masing sebesar 82,67% dan 86,92%. Selain itu, penerapan metode klasifikasi pada data ulasan pengguna Flip dengan rasio data latih dan data uji 80%:20% menghasilkan tingkat akurasi sebesar 76,68% menggunakan algoritma K-Nearest Neighbor [7].

Merujuk pada perumusan masalah yang telah dijelaskan sebelumnya, dalam penelitian ini penulis mengangkat judul "Implementasi Algoritma *K-Nearest Neighbors* untuk Klasifikasi Data Ulasan Pengguna Aplikasi Sulselbar *Mobile* Pada *Google Play Store*". Melalui penelitian ini, diharapkan proses klasifikasi ulasan pengguna dapat dilakukan secara akurat, sehingga dapat memberikan kontribusi signifikan dalam pengembangan system terutama dalam konteks layanan public seperti Bank.

Metode

Analisis sentimen, yang juga disebut penambangan opini, adalah cabang ilmu komputasi yang berfokus pada pengenalan dan pengkajian pendapat, emosi, sikap, evaluasi, penilaian, subjektivitas, maupun pandangan yang terkandung dalam suatu teks. Analisis sentimen bertujuan untuk menangkap dan mengambil informasi yang bersifat subjektif dari suatu teks, seperti pada ulasan produk, tanggapan pengguna, maupun isi dari media sosial [8]. Analisis sentimen dapat menjadi salah satu cara untuk mengetahui sejauh mana kepuasan pengguna. Melalui data yang tidak terstruktur, analisis ini mampu menghasilkan suatu kesimpulan yang bermakna [9].

Dalam analisis sentimen, dilakukan juga proses pembobotan dan klasifikasi teks yang bertujuan untuk mengevaluasi keterkaitan antara suatu kalimat dengan sejumlah dokumen menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF) [7]. TF-IDF adalah salah satu metode pembobotan fitur yang paling umum digunakan karena mampu memberikan tingkat akurasi dan recall yang cukup tinggi. Teknik ini sering diterapkan dalam bidang information retrieval. Pembobotan TF-IDF dianggap penting karena semakin sering suatu kata muncul dalam satu dokumen, semakin besar kontribusinya. Namun, jika kata tersebut sering muncul di banyak dokumen, maka kontribusinya justru menjadi lebih rendah [10]. *Term Frequency* (TF) dipandang merepresentasikan tingkat kepentingan suatu kata berdasarkan seberapa sering kata tersebut muncul dalam sebuah teks atau dokumen. Sementara itu, *Inverse Document Frequency* (IDF) adalah metode pembobotan token yang digunakan untuk mengukur seberapa sering suatu token muncul dalam keseluruhan kumpulan teks [11]. Untuk menghitung nilai TF-IDF, dapat diterapkan rumus pada persamaan (1-3) berikut [12].

$$tf = f_{t,d} \tag{1}$$

$$idf_d = log\left(\frac{N}{df_t}\right) \tag{2}$$

$$W_{(t,d)} = tf \times idf_d \tag{3}$$

Berdasarkan persamaan diatas, nilai *tf* menunjukkan frekuensi suatu term, sementara $f_{t,d}$ adalah jumlah kemunculan term *t* dalam dokumen *d*. Nilai idf_d merupakan *inverse document frequency*, yang mengukur seberapa jarang term tersebut muncul di seluruh dokumen. *N* menyatakan total jumlah dokumen, df_t adalah jumlah dokumen yang memuat term *t*, dan $w_{(t,d)}$ merepresentasikan bobot term *t* dalam dokumen *d* [12].

Metode K-Nearest Neighbor merupakan teknik klasifikasi yang bekerja dengan membandingkan jarak suatu objek terhadap data latih terdekat. Algoritma ini berfungsi untuk mengklasifikasikan objek berdasarkan tingkat kedekatannya dengan data pelatihan pada ruang fitur. *K-Nearest Neighbor* termasuk dalam metode pembelajaran berbasis contoh *(example-based learning)* atau *lazy learning,* serta tergolong dalam kategori instance-based learning [13]. Algoritma K-Nearest Neighbor (KNN) merupakan algoritma pembelajaran terawasi yang dapat diterapkan pada tugas klasifikasi maupun regresi. Algoritma ini bekerja dengan mencari *K* data terdekat dari sampel yang diberikan, kemudian menggunakan label kelas atau nilai dari data-data tersebut untuk memprediksi label atau nilai dari sampel tersebut [14]. Prinsip kerja K-Nearest Neighbor adalah menentukan jarak paling dekat antara data yang akan diklasifikasikan dengan k tetangga terdekatnya pada data uji [13].

1. Tentukan nilai parameter k. Nilai k dapat disesuaikan tergantung pada jumlah tetangga terdekat yang ingin dijadikan acuan dalam proses klasifikasi.
2. Hitung jarak antara data uji dengan seluruh data latih. Perhitungan ini dilakukan menggunakan rumus KNN, yaitu dengan mengukur tingkat kemiripan atau kedekatan antara data baru (data uji) dengan setiap data latih yang ada.
3. Urutkan hasil perhitungan jarak dari yang terkecil hingga terbesar (*ascending*).
4. Kumpulkan kategori Y, yaitu klasifikasi dari k tetangga terdekat berdasarkan nilai k. Kategori Y mewakili kelas atau objek data.
5. Prediksi kelas atau kategori objek dilakukan berdasarkan kategori tetangga terdekat yang jumlahnya paling banyak (mayoritas). Perhitungan jarak dilakukan menggunakan rumus *Euclidean Distance* berikut:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x-y)^2} \quad (4)$$

Dalam proses evaluasi model, penelitian ini menerapkan metode *confusion matrix*. *Confusion matrix*, yang juga disebut *error matrix*, adalah teknik yang umum digunakan dalam bidang *machine learning* dan statistika untuk menilai kinerja suatu model klasifikasi [15].

Tabel 1. *Confusion Metrix*[16]

	Predicted Negative	*Predicted Positif*
Actual Negative	TN	FP
Actual Positif	FN	TP

Keterangan:

TP : Data yang sebenarnya positif dan juga terklasifikasi sebagai positif.

FP : Data yang sebenarnya negatif namun salah diklasifikasikan sebagai positif.

FN : Data yang sebenarnya positif namun salah diklasifikasikan sebagai negatif.

TN : Data yang sebenarnya negatif dan juga terklasifikasi sebagai negatif.

1. *Accuracy* adalah nilai yang menunjukkan proporsi keseluruhan untuk menggambarkan tingkat akurasi dalam proses klasifikasi data dalam bentuk persen.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

2. *Precision* merupakan metode yang digunakan untuk menilai ketepatan model dalam menetapkan data ke kategori positif.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

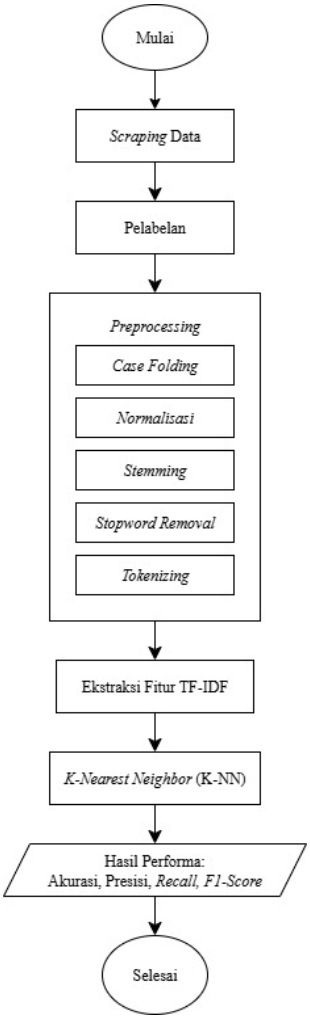
3. *Recall* adalah metode yang digunakan untuk menilai kemampuan model dalam mengklasifikasikan seluruh data positif dengan tepat.

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

4. *F1-Score* merupakan metrik untuk menyatukan nilai precision dan recall dalam satu ukuran. Metrik ini sangat bermanfaat saat data memiliki distribusi kelas yang tidak seimbang dalam proses klasifikasi.

$$F1 - Score = \frac{2\times(Precision\times Recall)}{Precision+Recall} \quad (8)$$

Perancangan



Gambar 1. Desain Penelitian

1. *Scraping* Data

Pengambilan data ulasan aplikasi Bank Sulselbar Mobile dari Google Play Store dilakukan dengan memanfaatkan library Google Play Scraper di Google Colab. Ulasan yang dikumpulkan berbahasa Indonesia dan berasal dari periode tahun 2024 hingga 2025, dengan total sebanyak 1000 data ulasan. Tetapi saat *scraping* dilakukan, data yang dikumpulkan berjumlah 2000 data agar memenuhi jumlah data positif dan negatif. Dimana jumlah data positif dan negatif yang akan digunakan dalam penelitian ini masing-masing berjumlah 500 data.

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
0	be408682-2f56-42e4-812c-e27611d9bf07	Surlaeni Amran	https://play-lh.googleusercontent.com/a-/ALV-U...	baik	5	0	2.22	2025-07-28 02:05:15	None	NaT	2.22
1	d035d124-83a4-4cc9-898a-cb47b0dc96dd	Amaar Nanci	https://play-lh.googleusercontent.com/a/ACg8oc...	sangat membantu meskipun kadang error	5	0	2.22	2025-07-24 01:54:20	None	NaT	2.22
2	c6fe3c20-a744-406d-a3c2-7b52cd6d3aba	Dion Dika	https://play-lh.googleusercontent.com/a/ACg8oc...	sangat membantu	5	0	2.22	2025-07-23 13:55:39	None	NaT	2.22
3	17ee0be6-4896-4a8b-a0dd-abff980c9845	Muh Fardhan	https://play-lh.googleusercontent.com/a/ACg8oc...	perbaiki masalah force close	3	0	None	2025-07-23 09:54:55	None	NaT	None
4	c4c1e242-2b73-4d11-9c40-eb46376e5d45	Nina Pia	https://play-lh.googleusercontent.com/a/ACg8oc...	gak bisa daftar BI fast gimana ni	3	0	None	2025-07-23 06:19:03	None	NaT	None

Gambar 2. Hasil *Scraping* Data

2. Pelabelan Data

Setelah data terkumpul, langkah berikutnya adalah memberikan label pada data tersebut. Proses pelabelan dilakukan menggunakan Google Colab. Usai pelabelan, dilakukan verifikasi manual untuk memastikan kualitas serta ketepatan label. Pengecekan ini dilakukan dengan memperhatikan jumlah bintang pada setiap ulasan. Selanjutnya, ulasan dikategorikan menjadi dua label: ulasan dengan skor bintang 5 sebagai label positif dan skor bintang 1 sebagai label negatif.

	userName	at	score	content	label
0	Surlaeni Amran	2025-07-28 02:05:15	5	baik	positif
1	Amaar Nanci	2025-07-24 01:54:20	5	sangat membantu meskipun kadang error	positif
2	Dion Dika	2025-07-23 13:55:39	5	sangat membantu	positif
5	Muh. Rifat Firjatullah Asir	2025-07-22 05:57:27	1	banyak BUG + ngelag	negatif
6	A Budiana, sh	2025-07-21 02:05:41	5	baik	positif

Gamar 3. Hasil Pelabelan Data

3. *Preprocessing*

Pada tahap *Preprocessing*, data yang telah diberi label akan diproses lebih lanjut melalui beberapa langkah, yaitu *case folding,* normalisasi, *stemming, stopword removal,* dan *tokenizing.*

a. Yang pertama yaitu proses *case folding* dimana pada proses ini dilakukan untuk mengubah semua teks menjadi huruf kecil.

Hasil Case Folding:

	content	casefolding	label
0	oke banget	oke banget	positif
1	Sangat mempermudah dalam segala hal yang saya ...	sangat mempermudah dalam segala hal yang saya ...	positif
2	Sangat mudah	sangat mudah	positif
3	Cabang Sengkang mantap.	cabang sengkang mantap.	positif

Gambar 4. Hasil Proses *Case Folding*

b. Yang kedua, normalisasi untuk mengubah kata tidak baku menjadi bentuk baku serta bentuk pada tahap ini juga digunakan kamus untuk meningkatkan hasil normalisasi [17].

Hasil Normalisasi:

	casefolding	normalized
0	oke banget	oke sekali
1	sangat mempermudah dalam segala hal yang saya ...	sangat mempermudah dalam segala hal yang saya ...
2	sangat mudah	sangat mudah
3	cabang sengkang mantap.	cabang sengkang mantap.

Gambar 5. Hasil Proses Normalisasi

c. Yang ketiga, *stemming* yaitu proses untuk mengubah kata ke bentuk dasarnya serta untuk meningkatkan hasil *stemming* yang dilakukan akan digunakan kamus yang dibuat manual [17].

Hasil Stemming:

	normalized	stemmed
0	oke sekali	oke sekali
1	sangat mempermudah dalam segala hal yang saya ...	sangat mudah dalam segala hal yang saya butuh
2	sangat mudah	sangat mudah
3	cabang sengkang mantap.	cabang sengkang mantap

Gambar 6. Hasil Proses *Stemming*

d. Selanjutnya, *stopword removal* yaitu tahapan membersihkan data teks agar hanya menyisakan kata-kata yang relevan dan bermakna bagi analisis berikutnya, juga pada tahap ini digunakan kamus agar hasilnya lebih optimal [17], untuk kamus yang digunakan peneliti mendapatkan dari github.

Hasil Stopword Removal:

	stemmed	stopword_removed
0	oke sekali	oke sekali
1	sangat mudah dalam segala hal yang saya butuh	sangat mudah segala butuh
2	sangat mudah	sangat mudah
3	cabang sengkang mantap	cabang sengkang mantap

Gambar 7. Hasil Proses *Stopword Removal*

e. Terakhir, dilakukan proses *tokenizing*, yakni tahap memisahkan kalimat menjadi bentuk kata-kata.

	stopword_removed	tokenized
0	oke sekali	[oke, sekali]
1	sangat mudah segala butuh	[sangat, mudah, segala, butuh]
2	sangat mudah	[sangat, mudah]
3	cabang sengkang mantap	[cabang, sengkang, mantap]

Gambar 8. Hasil Proses *Tokenizing*

4. Ekstraksi Fitur TF-IDF

Proses ini merupakan teknik dalam pengolahan bahasa alami (NLP) yang digunakan untuk mengubah teks menjadi bentuk numerik agar dapat dianalisis dan digunakan dalam model pembelajaran mesin. Hasil ekstraksi fitur menggunakan metode TF-IDF menunjukkan keberhasilan dalam mengenali kata-kata penting dari data teks. Tahapan ini sangat penting dalam pengembangan model analisis teks, karena pemilihan fitur yang tepat akan mempengaruhi kinerja model secara keseluruhan. Oleh karena itu, penting untuk melakukan evaluasi lebih lanjut guna memastikan bahwa istilah-istilah yang dipilih benar-benar relevan dan mewakili data secara akurat. Berikut adalah hasil ekstraksi fitur menggunakan TF-IDF.

	wow	xiomi	ya	yah	yahh	yakin	yang	yo	zaman
0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
1	0.000000	0.0	0.0	0.0	0.0	0.0	0.250829	0.0	0.0
2	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
3	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
4	0.314107	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0

Gambar 9. Hasil Proses TF-IDF

5. K-NN

Model ini dibangun menggunakan metode klasifikasi KNN dengan kernel linear dan parameter *K* sebanyak 3 tetangga terdekat. Data dibagi menjadi 80% untuk pengujian dan 20% untuk pelatihan, serta divalidasi menggunakan teknik cross-validation sebanyak 5 kali. Untuk mendukung proses klasifikasi, digunakan pembobotan TF-IDF dan pelabelan data berdasarkan referensi tertentu. Gabungan teknik-teknik ini ditujukan untuk meningkatkan tingkat akurasi dalam pengklasifikasian data.

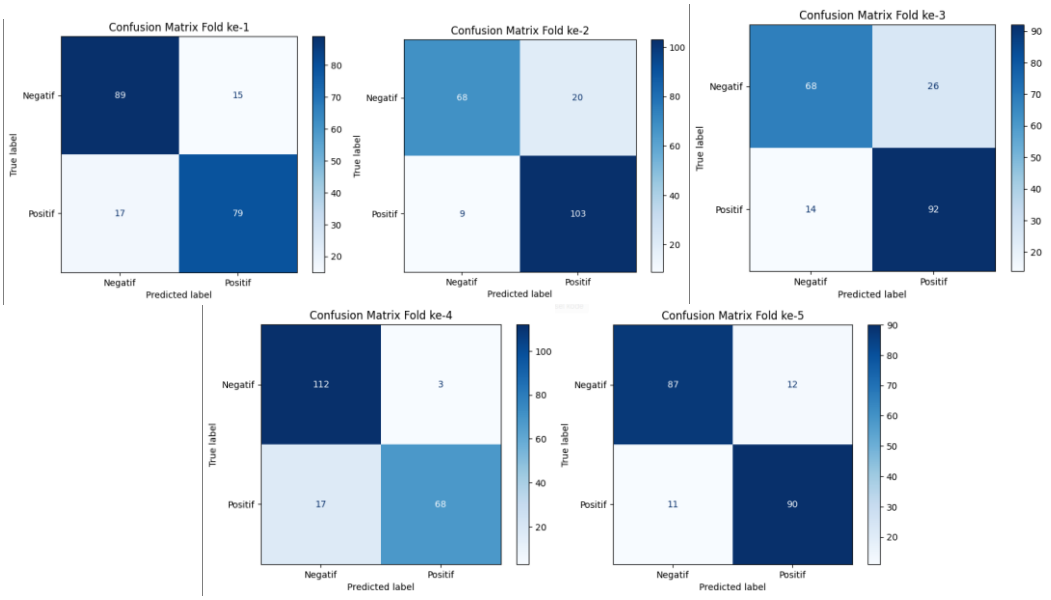
6. Hasil Performa

Tahap akhir dari proses ini adalah evaluasi, yang bertujuan untuk mengukur kinerja model KNN dengan mengacu pada tingkat akurasi. Evaluasi ini dilakukan untuk menilai sejauh mana keandalan algoritma dalam skenario pengujian yang telah dirancang. Melalui proses ini, diperoleh pemahaman yang lebih mendalam mengenai kemampuan KNN dalam memprediksi data yang belum pernah dilihat sebelumnya, sehingga dapat memberikan gambaran mengenai seberapa efektif algoritma ini jika diterapkan pada studi atau proyek lain di masa mendatang. Pada tahap ini, performa model KNN akan dianalisis menggunakan sejumlah metrik evaluasi, yaitu akurasi, presisi, *recall,* dan *f1-score.*

Perancangan mencakup analisis sistem dan desain sistem. Analisis sistem memaparkan identifikasi permasalahan pada sistem yang akan dirancang. Penjelasan ini dapat berisi solusi atau metode penyelesaian masalah, serta kebutuhan baru pada sistem yang akan dimodifikasi. Sementara itu, desain sistem meliputi perancangan input dan output, struktur atau hierarki sistem, prosedur pembacaan atau *flowchart* sistem, pembuatan prototipe sistem, serta arsitektur sistem.

Pemodelan

Pada pada penelitian ini, data ulasan pengguna terkait aplikasi Bank Sulselbar *Mobile* dikumpulkan melalui teknik *scraping* data menggunakan Google Colab. Data tersebut kemudian dilabelkan dengan label positif dan negatif, kemudian diproses melalui serangkaian tahap preprocessing (*case folding,* normalisasi, *stemming, stopword removal,* dan *tokenizing*). Selanjutnya diubah menjadi representasi numerik menggunakan teknik TF-IDF. Data tersebut kemudian dibagi menjadi 80:20 untuk data testing dan data uji, kemudian model dikembangkan menggunakan algoritma KNN dengan parameter optimal k = 3 serta menggunakan 5 *cross-validation.*



Gambar 10. *Confusion Matrix Fold* 1 Sampai 5

Pada Gambar 10 menunjukkan lima buah confusion matrix yang berasal dari proses evaluasi model klasifikasi menggunakan metode *cross-validation* dengan 5 fold. Setiap confusion matrix menggambarkan performa model dalam membedakan dua kelas, yaitu "Negatif" dan "Positif". Secara umum, model menunjukkan performa yang cukup konsisten di setiap fold, meskipun terdapat variasi kecil dalam akurasi dan kesalahan klasifikasi.

Pada Fold ke-1, model berhasil mengklasifikasikan mayoritas data dengan benar, namun masih terdapat 17 data positif yang salah diklasifikasikan sebagai negatif *(false negative),* serta 15 data negatif yang salah diklasifikasikan sebagai positif *(false positive).* Fold ke-2 menunjukkan kinerja yang sangat baik dalam mengenali data positif, dengan hanya 9 kesalahan *false negative,* namun masih terdapat 20 *false positive. Fold* ke-3 menunjukkan hasil yang sedikit lebih rendah karena adanya 26 *false positive* dan 14 *false negative.* Di sisi lain, Fold ke-4 menunjukkan kemampuan model yang sangat baik dalam mengenali data negatif dengan hanya 3 kesalahan *false positive,* meskipun masih terdapat 17 kesalahan false negative. Sementara itu, Fold ke-5 menampilkan kinerja yang seimbang, dengan 12 false positive dan 11 *false negative.* Secara keseluruhan, model mampu mengklasifikasikan data dengan cukup baik pada kelima fold. Meskipun masih terdapat kesalahan prediksi, terutama pada data positif yang diklasifikasikan sebagai negatif, hal ini menunjukkan bahwa model memiliki kecenderungan untuk lebih berhati-hati dalam memprediksi kelas positif. Dengan demikian, hasil dari *confusion matrix* ini dapat menjadi dasar untuk melakukan perbaikan model, seperti penyesuaian threshold atau pemilihan algoritma lain yang lebih sensitif terhadap kelas tertentu.

Tabel 1. Tabel Evaluasi Model KNN Fold 1 sampai 5

Fold	Akurasi	Presisi	Recall	F1 Score
1	84.0%	84.0%	84.0%	84.0%
2	85.5%	85.8%	85.5%	83.6%
3	80.0%	80.3%	80.0%	79.9%
4	90.0%	90.6%	90.0%	89.8%
5	88.5%	88.5%	88.5%	88.5%
Rata-rata	85.6%	85.4%	85.6%	85.5%

Berdasarkan Tabel 1, hasil evaluasi menunjukkan bahwa model KNN memiliki performa yang cukup baik dan konsisten di setiap fold. Fold ke-4 menunjukkan performa tertinggi, dengan akurasi 90.0%, presisi 90.6%, recall 90.0%, dan F1 score 89.8%. Ini menandakan bahwa pada fold tersebut, model sangat baik dalam mengenali kedua kelas (positif dan negatif) dengan tingkat kesalahan yang rendah. Sebaliknya, performa terendah terjadi pada Fold ke-3, dengan seluruh metrik hanya mencapai 80.0%–80.3%, dan F1 score 79.9%.

Hal ini menunjukkan bahwa pada fold tersebut, model memiliki lebih banyak kesalahan klasifikasi dibandingkan fold lainnya.

Jika dilihat dari rata-rata keseluruhan, model KNN memperoleh akurasi 85.6%, presisi 85.4%, recall 85.6%, dan F1 score 85.5%. Nilai yang seimbang antara presisi dan recall ini mengindikasikan bahwa model mampu mengenali kelas positif dan negatif dengan cukup baik, tanpa terlalu condong ke salah satu sisi (tidak terlalu banyak false positive maupun false negative).

Dengan performa yang cukup stabil di seluruh fold dan rata-rata metrik di atas 85%, dapat disimpulkan bahwa model KNN yang digunakan memiliki kualitas prediksi yang baik dan cukup andal dalam menyelesaikan tugas klasifikasi pada dataset yang diuji.

Kesimpulan

Penelitian ini berhasil mengimplementasikan algoritma K-Nearest Neighbors (KNN) untuk melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi Sulselbar Mobile yang diperoleh dari Google Play Store. Melalui proses pengumpulan data sebanyak 1000 ulasan, pelabelan, serta serangkaian tahapan preprocessing seperti case folding, normalisasi, stemming, stopword removal, dan tokenizing, data diubah menjadi representasi numerik menggunakan metode TF-IDF. Model KNN dengan parameter k = 3 kemudian diterapkan dan divalidasi menggunakan 5-fold cross-validation.

Hasil evaluasi performa model menunjukkan bahwa KNN mampu memberikan klasifikasi yang cukup baik dengan rata-rata akurasi sebesar 85,6%, presisi 85,4%, recall 85,6%, dan F1-score 85,5%. Fold ke-4 menunjukkan performa terbaik dengan akurasi 90% dan F1-score 89,8%, sementara fold ke-3 memiliki performa terendah dengan akurasi 80% dan F1-score 79,9%. Hal ini menunjukkan bahwa model memiliki kemampuan prediksi yang stabil dan dapat diandalkan dalam mengklasifikasikan ulasan menjadi sentimen positif dan negatif.

Dengan demikian, dapat disimpulkan bahwa penggunaan algoritma KNN yang dikombinasikan dengan teknik pembobotan TF-IDF efektif dalam menganalisis dan mengklasifikasikan sentimen ulasan pengguna aplikasi. Model ini dapat menjadi dasar dalam pengembangan sistem evaluasi layanan digital berbasis opini pengguna, khususnya untuk mendukung peningkatan kualitas layanan publik seperti aplikasi perbankan digital.

Daftar Pustaka

[1] U. Nur'aini, "Perbankan Syariah: Sebuah Pilar dalam Ekonomi Syariah," *Scholast. J. Pendidik. dan Kebud.*, vol. 4, no. 2, pp. 174–183, 2022.

[2] I. P. Trisela and U. Pristiana, "Analisis Perbandingan Kinerja Keuangan Bank Syariah dengan Bank Konvensional yang Terdaftar di Bursa Efek Indonesia Periode 2014-2018," *J. Ekon. Manaj.*, vol. 5, no. 2, pp. 83–106, 2020, doi: 10.56184/jam.v1i2.361.

[3] M. Muchran, Muchran, and Aenul, "Pengaruh Layanan Mobile Banking terhadap Loyalitas Nasabah Bank Sulselbar Cabang Utama Makassar," *SEIKO J. Manag. Bus.*, vol. 4, no. 3, pp. 27–31, 2022, doi: 10.37531/sejaman.vxix.357.

[4] Z. A. Zulkifly, N. Brasit, M. S. Alhaqqi, and S. Adelia, "Analisis Peningkatan Kualitas Layanan Mobile Banking dengan Pendekatan Metode E-Servqual," *JBMI (Jurnal Bisnis, Manajemen, dan Inform.*, vol. 19, no. 1, pp. 61–79, 2022, doi: 10.26487/jbmi.v19i1.21337.

[5] N. Anizah, Y. Salim, and L. B. Ilmawan, "Analisis Sentimen Terhadap Event Big Sale 11.11 Shopee di Media Sosial Instagram menggunakan Metode Naïve Bayes," *Bul. Sist. Inf. dan Teknol. Islam*, vol. 4, no. 1, pp. 25–34, 2023, doi: 10.33096/busiti.v4i1.1309.

[6] P. R. Prayoga, P. Purnawansyah, T. Hasanuddin, and H. Darwis, "Klasifikasi Daun Herbal Menggunakan K-Nearest Neighbor dan Support Vector Machine dengan Fitur Fourier Descriptor," *Edumatic J. Pendidik. Inform.*, vol. 7, no. 1, pp. 160–168, 2023, doi: 10.29408/edumatic.v7i1.17521.

[7] 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.

[8] Y. Ardian Pradana, I. Cholissodin, and D. Kurnianingtyas, "Analisis Sentimen Pemindahan Ibu Kota Indonesia pada Media Sosial Twitter menggunakan Metode LSTM dan Word2Vec," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 2389–2397, 2023.

[9] D. R. Alghifari, M. Edi, and L. Firmansyah, "Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia," *J. Manaj. Inform.*, vol. 12, no. 2, pp. 89–99, 2022, doi:

10.34010/jamika.v12i2.7764.
[10] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 422, 2021, doi: 10.30865/mib.v5i2.2845.
[11] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, "Film Review Sentiment Analysis Using TF-IDF and Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022.
[12] K. Pramayasa, I. M. D. Maysanjaya, and I. G. A. A. D. Indradewi, "Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE," *SINTECH (Science Inf. Technol. J.*, vol. 6, no. 2, pp. 89–98, 2023, doi: 10.31598/sintechjournal.v6i2.1372.
[13] M. A. Arsyad, T. Hasanuddin, and M. Hasnawi, "Implementasi K-Nearest Neighbor (KNN) pada Sistem Pakar Diagnosa Penyakit Untuk Menentukan Kelayakan Sapi sebagai Hewan Qurban Berbasis Web," *Bul. Sist. Inf. dan Teknol. Islam*, vol. 3, no. 3, pp. 167–173, 2022, doi: 10.33096/busiti.v3i3.632.
[14] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
[15] Muhammad Dzaki Arkaan Nasir and Syarif Hidayat, "Analisis Sentimen Ulasan Film Menggunakan Metode BiLSTM," *J. Inform. dan Teknol. Komput. (J-ICOM)*, vol. 5, no. 2, pp. 126–132, 2024, doi: 10.55377/j-icom.v5i2.8871.
[16] S. P. Backar, P. Purnawansyah, H. Darwis, and W. Astuti, "Hybrid Fourier Descriptor Naïve Bayes dan CNN pada Klasifikasi Daun Herbal," *J. Inform. J. Pengemb. IT*, vol. 8, no. 2, pp. 126–133, 2023, doi: 10.30591/jpit.v8i2.5186.
[17] N. Petty Wahyuningtyas, D. Eka Ratnawati, and N. Yudi Setiawan, "Root Cause Analysis (RCA) berbasis Sentimen menggunakan Metode K-Nearest Neighbor (K-NN) (Studi Kasus: Pengunjung Kolam Renang Brawijaya)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.* , vol. 7, no. 5, pp. 2515–2520, 2023, [Online]. Available: http://j-ptiik.ub.ac.id