ANALISIS SENTIMEN PENGGUNA PADA APLIKASI LINGOKIDS MENGGUNAKAN METODE K-NEAREST NEIGHBOUR

Indi Alpian Novansyah [1], Martanto [2], Tati Suprapti [3]
[1,3] Teknik Informatika, STMIK IKMI Cirebon
[2] Manajemen Informatika, STMIK IKMI Cirebon
Jalan Perjuangan, No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135
indialpian@gmail.com

ABSTRAK

Di era modern ini perkembangan teknologi dan komunikasi saat ini begitu pesat salah satunya adalah smartphone. Aplikasi pembelajaran anak-anak di smartphone dapat meningkatkan pengalaman belajar dengan desain yang menarik dan menyenangkan. Dalam konteks aplikasi pembelajaran anak-anak di smartphone, masih terdapat banyak kekurangan penelitian yang menyeluruh mengenai pengaruh desain anatarmuka pengguna terhadap efektivitas pembelajaran dan pengalaman pengguna. Dengan adanya pemahaman mengenai pengaruh desain antarmuka pengguna, pengembang aplikasi pembelajaran dan para pendidik dapat merancang aplikasi yang lebih efektif dan sesuai dengan kebutuhan dan karakteristik anak-anak. Perbaikan tampilan desain antarmuka menjadi lebih menarik, konsisten, dan memenuhi kepuasan pengguna. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap aplikasi Lingokids dengan menggunakan metode K-Nearest Neighbor. Teknik pengumpulan data yaitu dengan cara scraping data yang di konfigurasikan, kemudian diproses dengan menggunakan algoritma K-Nearest Neighbor untuk mengklasifikasikan sentimen pengguna positif, negatif, dan juga netral. Dalam penggunaan algoritma K-Nearest Neighbors (KNN) dengan dataset 80:20, terlihat peningkatan nilai akurasi seiring bertambahnya nilai K dari 1 (80,29%) menjadi 9 (87,65%). Meskipun nilai akurasi meningkat, evaluasi melalui confusion matrix menunjukkan performa yang bervariasi untuk setiap kelas. Kelas positif memiliki precision dan recall yang tinggi (88,96% dan 99,64%), sementara kelas negatif memiliki performa lebih rendah (precision 79,31% dan recall 46,94%). Hal ini menunjukkan bahwa meskipun akurasi meningkat dengan nilai K yang lebih tinggi, model KNN masih memiliki kesulitan dalam mengidentifikasi kelas negatif dengan tepat.

Kata kunci: *Teknologi, Aplikasi, Smartphone, K-Nearest Neighbor*

1. PENDAHULUAN

Di era modern ini perkembangan teknologi dan komunikasi saat ini begitu pesat salah satunya adalah smartphone. Smartphone telah menjadi bagian yang sangat penting bagi kehidupan manusia saat ini, termasuk pada kalangan anak-anak. Anak-anak sering menggunakan smartphone sebagai saran untuk bermain game, menonton video youtube, dan juga di gunakan untuk pembelajaran. Anak-anak usia dini hidup dalam dunia bermain, karena itulah dibutuhkan suatu media pembelajaran yang tidak hanya interaktif, namun juga menarik. Metode pembelajaran konvensional yaitu masih menggunakan buku dan poster dalam mengenalkan nama, belajar berhitung, dan lainnya yang cenderung monoton membuat minat belajar anak menurun. Sehingga membutuhkan metode pembelajaran yang baru agar dapat meningatkan semangat belajar anak.

Aplikasi pembelajaran untuk anak-anak di smartphone dapat meningkatkan pengalaman belajar dengan desain yang menarik dan menyenangkan. Dalam konteks aplikasi pembelajaran anak-anak di smartphone, masih terdapat banyak kekurangan penelitian yang menyeluruh mengenai pengaruh desain anatarmuka pengguna terhadap efektivitas pembelajaran dan pengalaman pengguna. Penelitian-penelitian yang ada umumnya masih terbatas dan belum memberikan pengalaman yang menyeluruh tentang bagaimana desain antarmuka pengguna dapat mempengaruhi pembelajaran anak-anak di smartphone. Pada usai 3 sampai 6 tahun ini anak-anak lebih cenderung dalam masa untuk bermain apalagi dizaman modern sekarang, anak-anak lebih memilih bermain game online maupun offline dikarenakan dari segi permainanya. Game online sendiri memiliki fitur yang menarik yang mendorong anak-anak tertarik bermain game sehingga tidak ada lagi waktu untuk belajar. Masalah lainnya adalah fitur dari aplikasi kids games yang agak sulit dimengerti karena kurang nya arahan langkah apa yang harus di lakukan selanjutnya sehinga membuat anak-anak bingung apa yang harus anak-anak lakukan sehingga membuat anak-anak cepat bosan.

Beberapa penelitian yang berkaitan telah dilakukakan sebelumnya, seperti dalam penelitian yang berjudul "Sentiment Analysis Of Reksadana On Bibit Applications Using The Naïve Bayes Method And K-Nearest Neighbor (KNN)" [1]. Adanya analisis sentimen mengenai bagaimana pendapat dari para pengguna aplikasi bibit reksadana menggunakan metode K-nearest neigbor (KNN) dan naïve bayes, dengan hasil scraping youtube sebanyak 33.292 dan scraping ulasan sebanyak 30.708 ulasan, lalu dilakukan tahap text processing dan labeling menggunakan library textlob dengan tingkat akurasi sebanyak 99%,99% dan naïve bayes sebanyak

99%,98% dapat disimpulkan bahwa sentiment netral lebih banyak dari sentimen positif dan sentimen positif lebih banyak dari sentimen negatif. Penelitian selanjutnya [2]yang berjudul "Analisis Sentimen Masyarakat Twitter Terhadap Piala Dunia U-20 2023 Di Indonesia Menggunakan Algoritma K-Nearest Neighbor". Dataset yang digukakan terdiri dari query dengan istilah 'U20HARUSJADI' dan 'KITAMAUMAINBOLA' di jejaring social twitter. Analisis dari 223 tweet menggungkapkan sentiment positif 72,84% dan sentiment negatif 27,16% sejak 1 maret sampai 30 april 2023. Hasil pengujian di peroleh dengan menggunakan nilai K=9 dengan akurasi 43% presisi 35% dengan recall 100%. Penelitian selanjutnya [3] yang berjudul "Analisis Big Data Sentimen Konsumen UMKM Sektor Kuliner Menggunakan Multi-Label K-Nearest Neighbor". Penelitian ini berfokus pada kasus multiclass classification dan pemilihan hyperparameter K tanpa teknik GridSearchCV terhadap hasil akurasi klasisfikasi.

Oleh karena itu, penelitian ini bertujuan untuk memberikan wawasan yang lebih mendalam pada aplikasi pembelajaran anak-anak di smartphone. Penelitian ini akan mengevaluasi dan menganalisis secara menyeluruh terhadap elemen-elemen desain antarmuka pengguna seperti tata letak, ikon, warna, dan fitur-fitur yang digunakan dalam aplikasi pembelajaran anak-anak di smartphone. Penelitian ini juga mengatasi permasalahan sentimen karena eberapa aplikasi kids games mungkin memiki masalah seperti kurangnya konten edukatif, ketidakcocokan usia, atau masalah teknis. Pendapat dan pengalaman pengguna dalam menggunakan aplikasi kid games sanngat penting untuk mengukur keberhasilan dan efektivitas aplikasi tersebut. Analisis sentimen pengguna dapat memberikan wawasan yang berharga tentang bagaimana orangtua dan anak menghadapi masalah ini.

Metode K-Nearest Neighbor (KNN) adlaah sebuah algoritma dalam machine learning yang digunakan untuk klasisfikasi dan regresi. K-NN bekerja dengan cara mencari sejumlah tetangga terdekat (K) dari sebuah titik data yang belum diklasifikasikan atau di prediksi, dan kemudian mengambil mayoritas kelas atau nilai target terdekat. Metode ini sederhanan dan juga mudah di impementasikan, tetapi sangat sensitive terdapat pemilihan parameter K. K-NN juga sangat cocok untuk data dengan karakteristik yang kurang kompleks dan dapat digunakan dalam berbagai bidang seperti pengenalan pola, pengolahan citra, dan pemroresan teks. Namun, performanya cenderung menurun jika dataset sanagt besar atau memiliki dimensi yang tinggi.

Implikasi dari adanya penelitian ini adalah, pengembang aplikasi dapat mengindentifikasi kelemahan dari aplikasi yang mereka buat dan memperbaiki untuk meningkatkan kualitas pendidikan dan hiburan kepada anak-anak. Pengembang aplikasi juga dapat merancang perangkat lunak yang lebih responsif terhadap kebutuhan pengguna dan meningkatkan interaktivitas untuk penngalaman belajar. Orang tua juga dapat memanfaatkan hasil penelitian ini, karena mereka dapat memilih aplikasi yang tepat untuk anak-anak mereka. Penelitian ini juga mendukung tentang pentingnya pemilihan konten yang pantas untuk anak-anak dalam dunia yang serba digital saat ini. Hasil penelitian ini juga dapat emmicu kolaborasi antar pengembang aplikasi, pendidik, dan orang tua untuk menciptakan standart baru dalam aplikasi pendidikan anak-anak. Selanjutnya membantu membangun literasi digital anak-anak sejak usia dini, mengjarkan mereka bagaimana menggunakan teknologi secara positif dan juga aman. Penelitian ini juga diharapkan dapat menjadi pedoman untuk membuka peluang lebih lanjut dalam di bidang ini.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen merupakan suatu bidang studi untuk menganalisis pendapat orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, dan topik. Analisis sentimen ini berfokus pada pendapat seseorang yang mengekspresikan atau menyiratkan sentimen positif, negatif, dan netral kebanyakan analis sentimen ini berkaitan dengan orang-orang di media sosial.[4]

2.2. K-Nearest Neighbors

Algoritma K-Nearest Neighbors adalah penentu klasifikasi berdasarkan contoh dasar yang tidak membangun, representasi deklaratif eksplisit kategori, tetapi bergantung pada label kategori yang melekat pada dokumen pelatihan mirip dengan dokumen tes. Algoritma K-Nearest Neighbors adalah sebuah metode klasifikasi terhadap objek berdasarkan data training yang jaraknya paling dekat. K-Nearest Neighbors memiliki prinsip sederhana, bekerja berdasarkan jarak terpendek dari sampel uji ke sampel latih. Menurut Safitri.[5] algoritma k-nearest neighbors memiliki tahapan:

1. Menentukan parameter k (jumlah tetangga paling dekat).
2. Tentukan bobot untuk setiap term dengan menggunakan Term Weighting TF-IDF.
3. Hitung kemiripan antar dokumen dengan menggunakan cosine similarity.
4. Urutkan hasil perhitungan cosine similarity dari besar ke kecil.
5. Ambil sebanyak K yang paling tinggi kemiripannya dengan dokumen yang diklasifikasikan, tentukan kelasnya.

2.3. Confusion Matrix

Confusion matrix adalah metode yang digunakan untuk perhitungan akurasi, nilai recall, precision.[5]

Tabel 1. Model confusion matrix

	Pred.positif	**Pred.negatif**
True positif	TP	FN
True negatif	FN	TN

Dari Tabel 1 dapat di lihat bahwa :

a. TP (True Positive) = Jumlah prediksi yang benar dari data yang relevant.
b. FP (False Positive) = Jumlah prediksi yang salah dari data yang tidak relevant.
c. FN (False Negative) = Jumlah prediksi yang salah dari data yang tidak relevant.
d. TN (True Negative) = Jumlah prediksi yang benar dari data yang relevant.

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$F1\ score = \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

$$Accurasy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

2.4. Literature Review

Pada penelitian yang berjudul Klasifikasi Komentar pada Pembelajaran E-Learning menggunakan Analisis Sentimen dengan Metode K-Nearest Neighbour yang dilakukan oleh Prasetyo Maulidina & Abdurrachman Bachtiar, pada tahun 2022. Pembelajaran interaktif merupakan salah satu metode pembelajaran yang diminati terutama disaat pandemic seperti hari ini. Salah satu dari metode pembelaran paling interaktif yang perkembangannya sangat cepat adalah pembelajran interaktif berbasis e-learning. Dengan demikian penelitian ini mencoba mengklasifikasikan komentar pada pembelajaran e-learning. Proses pengklasifikasian dilakukakn menggunakan metode K-Nearest Neighbour. Pada proses analisis sentimen ini terdapat 4 proses utama yaitu Text Processing, Term Weighting (TF IDF), penghitungan Cosine Similarity, dan terakhir klasifikasi menggunakan metode K_nearest Neighbour. Dari hasil pengujian, diketahui bahwa hasil akurasi yang terbaik sebesar 87,71% diperoleh sat nilai K=1, dan 71,42% menggunakan nilai K=2.[6]

Pada penelitian yang berjudul Analisis Sentimen Pada Ulasan Aplikasi Generative Pre-Trained Transformer GPT Menggunakan Metode K-Nearest Neighbour (KNN) yang dilakukan oleh Fahriza & Riza, pada tahun 2023. Teknologi chatbot semakin populer dan aplikasi yang menggunakan model generatif seperti GPT seakin banyak digunakan. GPT merupakan model Bahasa yang dilatih dengan data yang besar untuk menghasilkan teks yang mirip dengan teks manusia. Namun, penting untuk memahami sentimen atau pendapat pengguna terhadap aplikasi chat generatif ini. Penelitian ini mendapatkan hasil dengan nilai K=9 dengan akurasi 0.966, presisi 0.483, dan recall 0.5 dan terakhir F1 score:0.491. Dalam hasil ini akurasi model menunjukan 0.966 atau sekitar 96.6% dengan sentimen positif 83.64%, sentimen negatif 3.48%, dan netral 3.27%.[7]

Selanjutnya penelitian yang berjudul Analisis Sentimen Pengguna Transportasi Online Maxim Pada Instagram Menggunakan Naïve Bayes Classifier dan K-Nearest Neighbor yang dilakukan oleh Warraihan pada tahun 2023. Transportasi online adalah bentuk transportasi berbasis internet yang mencakup semua aspek proses transaksi, termasuk pemesanan, pelacakan rute, pembayaran, dan penilaian layanan terhadap transportasi online tersebut. Penelitian ini melakukan analisis sentimen terhadap pendapat/opini pengguna maxim di Instagram menggunakan algoritma Naïve Bayes Classfier (NBC) dan K-Nearest Neighbour (KNN). Hasil accuracy pada data sentimen menggunakan algoritma NBC yaitu 81,03% dan pada algoritma KNN dengan nilai K=3 yaitu sebesar 80,72%.[8]

Penelitian selanjutnya yang berjudul Analisis Sentimen Terhadap Penerapan Sistem Ganjil Genap Menggunakan Metode K - Nearest Neighbor (KNN) yang dilakukan oleh Rohmansyah, pada tahun 2022. Penerapan system ganjil genap pada rambu lalu lintas merupakan kebijakan pemerintah terutama untuk mengurangi kemacetan di ibu kota. Dalam penelitian ini mencoba untuk menganalisis hasil sentimen pada platform media sosial untuk kebijakan ganjil genap tersebut. Pada penelitian ini menggunakan metode K-Nearest Neighbour dan mendapatkan akurasi 75,36% dengan jumlah data sebanyak 350 data dan menghasilkan 128 sentimen positif dan 222 sentimen negatif.[9]

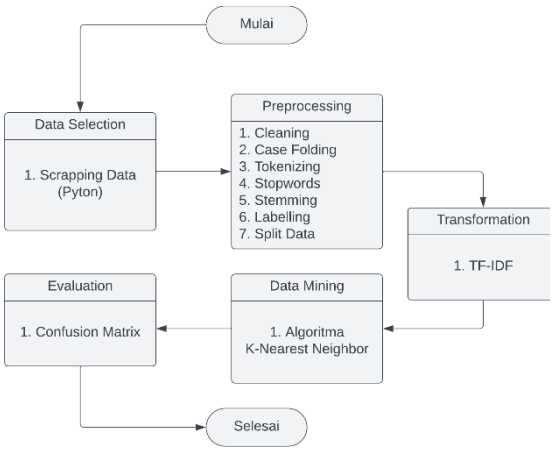
Penelitian selanjutnya yang berjudul Analisis Sentimen Terhadap Layanan Internet Di Indonesia Menggunakan Metode K-Nearest Neighbor yang dilakukan oleh Saputra & Pribadi, pada tahun 2023. Analisis sentimen terhadap lambatnya internet di Indonesia ini sangat penting untuk masyarakat Indonesia. Tanggapan klasifikasi menjadi sentimen positif dan negatif menggunakan metode k-Nearest Neighbor. Dari 378 opini pengguna youtube yan berhasil di kumpulkan, didapat 121 opini positif dan 253 opini negatif. Selanjutnya dioptimasi dengan rapidminer. Sehingga mendapatkan akurasi 75,42% sedangkan menggunakan naïve bayes 60,71% dapat disimpulkan bahwa metode K-Nearest Neighbor tingkat akurasi lebih tinggi di bandingkan Naïve Bayes.[10]

Penelitian selanjutnya yang berjudul Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor yang dilakukakan oleh Septian, pada tahun 2019. Persepakbolaan Indonesia belakangan ini memiliki banyak polemik mulai dari kasus pengaturan skor, pergantian pelatih timnas senior hingga pergantian ketua umum

Persatuan Sepak bola Seluruh Indonesia (PSSI). Tujuan dari penelitian ini adalah untuk menganalisis sentimen pada setiap kalimat dari pengguna twitter terhadap persepakbolaan Indonesia apakah memiliki sentimen negatif atau positif menggunakan K-Nearest Neighbor. Pembobotan kata menggunakan Term Frequency-Invers Document Frequency (TF-IDF). Pada tahap validasi data dilakukan pengujian silang sebanyak 10 kali menggunakan k-fold cross validation, kemudian diklasifikasikan dengan metode K-Nearest Neighbor dapat menghasilkan akurasi yang cukup baik. Dari 2000 data tweet berbahasa indonesia didapatkan hasil akurasi optimal pada nilai k=23 sejumlah 79.99%.[4]

Penelitian selanjutnya yang berjudul Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter yang dilakukan oleh Nasution & Hayaty, pada tahun 2019. Masalah yang dibahas dalam penelitian ini adalah perbandingan kinerja dua algoritma klasifikasi untuk analisis sentimen, yaitu K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM). Analisis sentimen adalah proses mengidentifikasi sentimen positif, negatif, atau netral dari teks. Terdapat dua jenis metode pembelajaran mesin yang dapat digunakan untuk analisis sentimen: supervised learning dan unsupervised learning. Hasil pada perhitungan akurasi menunjukkan bahwa metode Support Vector Machine lebih unggul dengan nilai 89,70% tanpa K-Fold Cross Validation dan 88,76% dengan K-Fold Cross Validation. Sedangkan pada perhitungan waktu proses metode K-Nearest Neighbor lebih unggul dengan waktu proses 0.0160s tanpa K-Fold Cross Validation dan 0.1505s dengan K-Fold Cross Validation.[11]

3. METODE PENELITIAN



Gambar 1. Tahapan penelitian

Pada penelitian ini terdapat beberapa tahapan seperti yang ditunjukan pada gambar 1 sebagai berikut:

a. Data Selection
   Pada tahapan ini penelitian ini melakukan pengambilan data ulasan pada playstore 30 Oktober 2023. Sebelum melakukan pengambilan data peneliti harus mengkonfigurasi pada googgle collab yang menggunakan Bahasa pyton agar pengambilan data bias dilakukan. Peneliti juga memilih data mana saja yang di ambil seperti rating, nama, tanggal, dan yang paling terping ulasan. Data yang diambil hanya data ulasan yang terdapat pada aplikasi Lingokids.
b. Preprocesing
   Dari data yang sudah di ambil, setelah itu data akan masuk tahap preprocessing. Berikut adalah tahapan preprocessing:
   1. Cleaning merupakan menghapus data duplikat, data yang tidak relevan, atau data yang mengganggu, seperti noise
   2. Case Folding merupakan proses untuk mengubah semua data menjadi huruf kecil.
   3. Tokenizing merupakan proses pemecahan sekumpulan karakter dalam suatu teks kedalam satuan kata, pada proses ini juga dapat menghilangkan karakter pembatas, menghapus angka, dan karakter yang bukan huruf.
   4. Stopwords merupakan proses menghilangkan kata yang tidak miliki makna atau tidak penting menggunakan stopword list seperti kok, lah, dll.
   5. Stemming merupakan proses pencarian kata dasar dari tiap kata. Pada penelitian ini peneliti menggunakan sebuah file yang berisi imbuhan yang bertujuan untuk menghilangkan imbuhan untuk mendapatkan akar kata.
   6. Labeling merupakan proses menandai atau memberikan klasifikasi tertentu kepada teks atau ulasan pengguna berdasarkan sentimen.
   7. Split data merupakan proses membagi data set seperti data training (data latih) dan data testing (data uji).
c. Transformation
   Setelah selesai tahap preprocesing selanjutnya, melakukan tahap transformation dengan menggunakan Term Weighting (TF-IDF) merupakan proses untuk memberikan pembobotan kata dengan mencari nilai Term Frequency (TF), kemudian mencari nilai Document Frequency (DF), lalu mencari nilai Invers Document Frequency (IDF) setelah itu baru menghitung bobot.
d. Data Mining
   Pada tahapan ini akan dilakukan pemodelan terhadap data ulasan yang sudah dilakukan preprocessing. Kemudian pada tahapan pemodelan ini peneliti menggunakan algoritma K-Nearest Neighbor untuk mendapatkan nilai sentimen positif dan sentimen negatif dari data ulasan.
e. Evaluation
   Setelah semua tahap selesai, kemudian melakukan proses evaluasi. Pada tahapan ini akan dilakukan evaluasi metode klasifikasi dengan mengukur performa menggunakan confusion matrix terhadap algoritma K-Nearest Neighbor.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Penelitian ini menggunakan teknik scrapper data dengan menggunakan library 'google-play-scraper untuk mengambil data ulasan dari aplikasi Lingokids yang didukung menggunakan Bahasa pemograman python yang kemudian di eksekusi menggunakan Google Collab. Data yang di ambil berupa informasi seperti username, waktu, rating ulasan, dan isi konten ulasan. Peneliti mengambil 1761 review teratas yang di urutkan berdasarkan kaitan dengan topik penelitian. Dengan Teknik scrapping data ini, peneliti dapat dengan mudah mengumpulkan data ulasan yang cukup besar dari Google Play Store.

4.2. Preprocesing

Pada tahapan ini mencakup beberapa teknik. Berikut ini adalah langkah-langkah yang dilakukan selama data preprocessing:

1. Cleaning merupakan menghapus data duplikat, data yang tidak relevan, atau data yang mengganggu, seperti noise.
2. Case folding merupakan proses untuk mengubah semua data menjadi huruf kecil atau huruf besar.
3. Tokenizing merupakan proses memisahkan teks menjadi unit-unit kecil atau disebut dengan token.
4. Stopword merupakan proses untuk menghilangkan kata yang tidak memiliki makna.
5. Stemming merupakan proses menghapus imbuhan kata sehingga kata-kata tersebut dapat diidentifikasi sebagai bentuk dasar yang sama.
6. Labeling merupakan proses yang dilakukan untuk menentukan nilai sentimen berdasarkan nilai score.
7. Split data merupakan proses membagi data menjadi data latih dan data uji.

4.3. Transformation

Transformasi data adalah proses untuk mengubah atau memanipulasi data dari satu bentuk ke bentuk yang berbeda. Pada transfromasi data ini peneliti menggunakan salah satu ekstraksi fitur untuk tranformasi data yaitu TF-IDF. TF-IDF adalah satu-satunya transformasi yang sering digunakan dalam kasus analisis sentimen yang berfungsi mengubah teks menjadi bilangan numerik.

4.4. Data Mining

Data mining adalah proses analisis besar data set unutk mengidentifikasi pola atau infomasi yang bermanfaat dan bermakna. Pada tahap ini dilakukan penerapan algoritma K-Nearest Neighbor untuk melakukan analisis sentimen pada data ulasan pengguna yang ada pada aplikasi Lingokids. Algoritma K-Nearest Neighbors (KNN) sering digunakan dalam analisis sentimen karena sifatnya yang relatif mudah dipahami, mudah diimplementasikan, dan dapat memberikan hasil yang cukup baik dalam beberapa kasus. Data yang akan digunakan adalah data yang telah melalui tahapan seperti preprocessing dan transformasi. Pada tahapan ini maka data akan di split jadi 3 model yaitu 60:40, 70:30, dan 80:20. Data akan dilatih dengan tetangga terdekat dengan nilai K=1, K=3, K=5, K=7, dan K=9.

Tabel 2. Hasil akurasi dengan split data

Perbandingan	Akurasi Algoritma KNN				
	K=1	K=3	K=5	K=7	K=9
60:40	80,15 %	80,15%	81,32%	82,21%	84,85%
70:30	76,86%	80,00%	80,20%	80,98%	84,90%
80:20	80,29%	79,71%	81,76%	87,06%	87,65%

Dari Tabel 2. maka dapat disimpulkan nilai akurasi akurasi meningkat seiring dengan bertambahnya nilai K. dalam kasus lain dari Tabel 2 nilai akurasi relatif konsisten seperti pada contoh model split 60:40, nilai akurasi relative stabil di sekitar 80-85% ketika K berubah. Dapat dilihat juga akurasi tertinggi yang dihasilkan pada penelitian ini berada pada perbandingan 80:20 dengan nilai K=9 dan akurasi sebesar 87,65%.

4.5. Evaluation

Penelitian ini akan melakukan pengujian dengan menggunakan confusion matrix yang diperoleh dari hasil pembagian split data 3 model dengan menggunakan algoritma k-nearest neighbors dimana pada Tabel 2. Akurasi dengan split data 80:20 lebih tinggi dibandingkan model lainnya.

Tabel 3. Hasil confusion matrix

Kelas	Pred. Negatif	Pred. Netral	Pred. Positif
True Negatif	23	2	24
True Netral	5	0	10
True Positif	1	0	275

Tabel 3. adalah hasil evaluasi model klasifikasi. Dalam konteks ini, terdapat tiga kelas yang diprediksi oleh model: "Negative", "Neutral", dan "Positive". Setiap sel di dalam confusion matrix mencerminkan jumlah contoh dari data uji yang termasuk dalam kombinasi tertentu dari prediksi yang dilakukan oleh model dan nilai sebenarnya dari kelasnya. Peneliti akan menjelaskan lebih detail mengenai confusion matrix lebih detail:

a. Dari 49 contoh yang sebenarnya merupakan kelas "Negative" model memprediksi 23 contoh sebagai "Negative" yang benar (true negative), 2 contoh sebagai "Neutral" yang seharusnya "Negative" (false negative) 24 contoh sebagai "Positive" yang seharusnya "Negative" (false negative).
b. Dari 15 contoh yang sebenarnya merupakan kelas "Neutral" model memprediksi 5 contoh sebagai "Negative" yang seharusnya "Neutral" (false positive), 0 contoh sebagai "Neutral" yang

benar (true negative), 10 contoh sebagai "Positive" yang seharusnya "Neutral" (false negative).

c. Dari 276 contoh yang sebenarnya merupakan kelas "Positive" model memprediksi 1 contoh sebagai "Negative" yang seharusnya "Positive" (false positive), 0 contoh sebagai "Neutral" yang seharusnya "Positive" (false positive), 275 contoh sebagai "Positive" yang benar (true positive).

Dengan menggunakan confusion matrix ini, kita dapat menghitung berbagai metrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk mengevaluasi kinerja model terhadap secara keseluruhan. Berikut perhitungan dari hasil confusion matrix dari Tabel 2.

a. Kelas negatif

$$Precision = \frac{TP}{TP + FN} = \frac{23}{23 + 5 + 1} = \frac{23}{29} = 0{,}7931 = 79{,}31\% \quad (5)$$

$$Recall = \frac{TP}{TP + FN} = \frac{23}{23 + 2 + 24} = \frac{23}{49} = 0{,}4694 = 46{,}94\% \quad (6)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0{,}7931 \times 0{,}4694}{0{,}7931 + 0{,}4694} = 0{,}5914 = 59{,}14\% \quad (7)$$

b. Kelas positif

$$Precision = \frac{TP}{TP + FN} = \frac{23275}{275 + 24 + 10} = \frac{275}{309} = 0{,}8896 = 88{,}96\% \quad (8)$$

$$Recall = \frac{TP}{TP + FN} = \frac{275}{275 + 1 + 0} = \frac{275}{276} = 0{,}9964 = 99{,}44\% \quad (9)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0{,}8896 \times 0{,}9964}{0{,}8896 + 0{,}9964} = 0{,}9401 = 94{,}01\% \quad (10)$$

c. Akurasi keseluruhan

Total True Predictions = TP Negative + TP Neutral + TP Positive = 23 + 0 + 275 = 298

Total Predictions = Total jumlah elemen dalam confusion matrix = 23 + 2 + 24 + 5 + 0 + 10 + 1 + 0 + 275 = 340

$$Accurasy = \frac{Total\ true\ prediction}{Total\ prediction} = \frac{298}{340} = 0{,}8765 = 87{,}65\% \quad (11)$$

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dapat dimpulkan bahwa Nilai akurasi cenderung meningkat seiring bertambahnya nilai K pada algortima K-Nearest Neighbor dapat dilihat pada model split data 80:20 dengan menggunakan nilai K=1 mendapatkan akurasi 80,29% sedangkan jika menggunakan nilai K=9 mendapatkan akurasi 87,65%, dengan semakin menambahkan jumlah tetangga terdekat berarti meningkat juga performa model. Melalui confusion matrix, terlihat bahwa performa model tergantung pada kemampuan memprediksi setiap kelas dengan benar seperti, kelas positif memiliki precision dan recall yang tinggi yaitu sebesar 88,96% dan 99,64%, sedangkan kelas negatif menunjukan perfoma yang lebih rendah dengan precision 79,31% dan recall 46,94%. Beberaa saran yang dapat dilakukan untuk penelitian selanjutnya seperti. Melakukan penelitian lebih lanjut mengenai pengoptimalan model, dapat dilihat performa model kelas negatif yang dapat diperbaiki seperti pengoptimalan fitur atau teknik preprocessing yang lebih efektif lagi, membandingkan kinerja model K-Nearest Neighbor dengan metode klasifikasi lainnya seperti Naïve bayes dan SVM, dan Melakukan pengujian model pada data set yang lebih luas.

DAFTAR PUSTAKA

[1] A. Fitriyani and A. Triayudi, "Sentiment Analysis of Reksadana on Bibit Applications Using the Naïve Bayes Method and K-Nearest Neighbor (Knn)," *J. Ris. Inform.*, vol. 4, no. 2, pp. 127–134, 2022, doi: 10.34288/jri.v4i2.304.

[2] S. A. Ramdhani and S. Waluyo, "ANALISIS SENTIMEN MASYARAKAT TWITTER TERHADAP PIALA DUNIA U-20 2023 DI INDONESIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR ANALYSIS OF PUBLIC SENTIMENT ON TWITTER SOCIAL MEDIA FOR THE 2023 U20 WORLD CUP IN INDONESIA USING THE K- NEAREST NEIGHBOR (KNN) AL," vol. 2, no. September, 2023.

[3] M. R. Fauzi, R. A. Pratama, P. Laksono, and P. Eosina, "Penerapan Big Data Menggunakan Algoritma Multi-Label K-Nearest Neighbor dalam Analisis Sentimen Konsumen UMKM Sektor Kuliner," *Krea-TIF*, vol. 9, no. 1, p. 9, 2021, doi: 10.32832/kreatif.v9i1.3587.

[4] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-

Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.

[5] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.

[6] H. Prasetyo Maulidina and F. Abdurrachman Bachtiar, "Klasifikasi Komentar pada Pembelajaran E-Learning menggunakan Analisis Sentimen dengan Metode K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 5, pp. 2301–2307, 2022, [Online]. Available: http://j-ptiik.ub.ac.id

[7] M. N. Fahriza and N. Riza, "ANALISIS SENTIMEN PADA ULASAN APLIKASI CHAT GENERATIVE PRE-TRAINED TRANSFORMER GPT MENGGUNAKAN METODE KLASIFIKASI K-NEAREST NEIGHBOR(KNN) Sistematic Literature Review," *J. Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1351–1358, 2023.

[8] D. A. Warraihan, I. Permana, Mustakim, and R. Novita, "Analisis Sentimen Pengguna Transportasi Online Maxim Pada Instagram Menggunakan Naïve Bayes Classifier dan K-Nearest Neighbour," vol. 7, no. 3, pp. 1134–1143, 2023, doi: 10.30865/mib.v7i3.6336.

[9] F. A. Rohmansyah, B. Bintoro, and I. Santoso, "Analisis Sentimen Terhadap Penerapan Sistem Ganjil Genap Menggunakan Metode K-Nearest Neighbor (Knn)," *iKRAITH-INFORMATIKA*, vol. 7, no. 2, pp. 165–169, 2022.

[10] D. Saputra and M. R. Pribadi, "Analisis Sentimen Masyarakat terhadap Layanan Provider Internet di Indonesia menggunakan SVM," *MDP Student Conf.*, vol. 2, no. 1, pp. 32–41, 2023, doi: 10.35957/mdp-sc.v2i1.4554.

[11] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.

.