

## IMPLEMENTASI ALGORITME *K-NEAREST NEIGHBOUR* DAN *LEXICON BASED* UNTUK ANALISIS SENTIMEN KEPUASAN PENGGUNA APLIKASI GRAMEDIA DIGITAL PADA MEDIA SOSIAL TWITTER

Al Adiat Firman Alamsyah<sup>1</sup>, Sri Mulyati<sup>2\*</sup>

<sup>1</sup>Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

<sup>2\*</sup>Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: <sup>1</sup>1911510012@student.budiluhur.ac.id, <sup>2\*</sup>sri.mulyati@budiluhur.ac.id  
(\* : corresponding author)

**Abstrak-**Opini atau pandangan yang diungkapkan oleh masyarakat menjadi salah satu tolok ukur yang digunakan untuk mengevaluasi aspek tertentu seperti produk atau layanan. Dalam konteks aplikasi Gramedia Digital yang memiliki penggunaan luas untuk membeli dan membaca buku secara digital, penting untuk mengkaji dan menganalisis sentimen yang terkait untuk memahami bagaimana pendapat pengguna terhadap aplikasi tersebut. Dengan memanfaatkan *platform* media sosial seperti Twitter, kita dapat mengumpulkan berbagai sentimen yang muncul dari masyarakat terkait aplikasi Gramedia Digital. Analisis ini bertujuan untuk mengklasifikasikan sentimen tersebut menjadi positif, negatif, atau netral. Salah satu metode yang berguna dalam mengumpulkan informasi dan menganalisis pandangan masyarakat adalah melalui teknik analisis sentimen. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi Gramedia Digital menggunakan metode *K-Nearest Neighbour* (KNN) dan *Lexicon Based*. Tujuan lainnya adalah untuk memperoleh pemahaman yang lebih mendalam mengenai kepuasan pengguna terhadap aplikasi tersebut. Metode *K-Nearest Neighbour* (KNN) dan *Lexicon Based* digunakan dalam kombinasi untuk menganalisis sentimen pengguna terhadap aplikasi Gramedia Digital. Dengan menggunakan KNN, teks tweet yang mengandung opini dan sentimen diidentifikasi dan diklasifikasikan menjadi dua kategori, yaitu sentimen positif dan negatif. Metode *Lexicon Based* digunakan untuk memperkaya pemahaman sentimen dengan mengacu pada daftar kata-kata yang dikategorikan sebagai positif, negatif, atau netral. Kombinasi kedua metode ini diharapkan dapat mengoptimalkan hasil identifikasi sentimen pada teks *tweet* yang diambil dari media sosial Twitter. Hasil analisis sentimen menggunakan metode KNN dan *Lexicon Based* pada 1330 *tweet* dari media sosial Twitter dengan kata kunci 'Gramedia Digital' dan 'Gram dig' menunjukkan adanya dua kategori sentimen, yaitu positif dan negatif. Dalam *dataset* yang digunakan, sebanyak 1330 *tweet* dari periode 1 Januari 2023 hingga 15 Mei 2023, diperoleh akurasi tertinggi sebesar 75,97% dengan menggunakan nilai  $k = 3$ , dan rasio 60% data latih dan 40% data uji.

**Kata Kunci:** Analisis sentimen, *K-Nearest Neighbour*, *Lexicon*, Gramedia Digital, Twitter

## IMPLEMENTATION OF *K-NEAREST NEIGHBOUR* AND *LEXICON BASED* ALGORITHMS FOR SENTIMENT ANALYSIS OF USER SATISFACTION GRAMEDIA DIGITAL APPLICATION ON TWITTER SOCIAL MEDIA

**Abstract-** Opinions or views expressed by the public are one of the benchmarks used to evaluate certain aspects such as products or services. In the context of Gramedia Digital's applications that have widespread use for buying and reading books digitally, it is important to examine and analyze the sentiments associated to understand how users perceive the application. By utilizing social media platforms such as Twitter, we can collect various sentiments that arise from the public regarding the Gramedia Digital application. This analysis aims to classify these sentiments into positive, negative, or neutral. One method that is useful in gathering information and analyzing public views is through sentiment analysis techniques. This study aims to analyze user sentiment towards the Gramedia Digital application using the *K-Nearest Neighbor* (KNN) and *Lexicon Based* methods. Another goal is to gain a deeper understanding of user satisfaction with the application. The *K-Nearest Neighbor* (KNN) and *Lexicon Based* methods are used in combination to analyze user sentiment towards the Gramedia Digital application. By using KNN, tweet texts that contain opinions and sentiments are identified and classified into two categories, namely positive and negative sentiments. The *Lexicon Based* method is used to enrich the understanding of sentiment by referring to a list of words that are categorized as positive, negative, or neutral. The combination of these two methods is expected to optimize the results of sentiment identification in tweet texts taken from social media Twitter. The results of sentiment analysis using the KNN and *Lexicon Based* methods on 1330 tweets from social media Twitter with the keywords 'Gramedia Digital' and 'Gramdig' show that there are two categories of sentiment, namely positive and negative. In the dataset

used, a total of 1330 tweets from the period January 1 2023 to May 15 2023, obtained the highest accuracy of 75.97% using a value of  $k = 3$ , and a ratio of 60% training data and 40% test data.

**Keywords:** Sentiment analysis, K-Nearest Neighbors, Lexicon, Gramedia Digital, Twitter

## 1. PENDAHULUAN

Analisis sentimen adalah proses pengumpulan, pengolahan, dan analisis data untuk menentukan sentimen atau opini yang terkandung dalam teks atau data yang dianalisis. Analisis sentimen dapat dilakukan pada berbagai jenis data, termasuk teks di media sosial, ulasan produk atau ulasan aplikasi.

Gramedia digital adalah sebuah aplikasi atau *platform* yang dikembangkan oleh PT. Gramedia Digital Nusantara, sebuah perusahaan yang bergerak di bidang media cetak dan transformasi digital di Indonesia. Gramedia Digital menawarkan buku-buku digital, majalah, dan bahan bacaan lain yang dapat diakses melalui aplikasi tersebut.

Data yang diperoleh dari Twitter, disebutkan bahwa Indonesia menduduki peringkat 5 pengguna terbanyak di dunia [1]. Suatu kesimpulan bisa didapatkan jika dilakukan penelitian lebih lanjut pada cuitan opini yang berisi sentimen. Berdasarkan hal tersebut, maka Twitter adalah sumber data yang dapat digunakan untuk dilakukan penelitian analisis sentimen. Kepuasan pengguna aplikasi Gramedia Digital yang menggunakan Twitter sebagai sarana untuk mengutarakan opini sangat penting. Opini yang dikumpulkan melalui Twitter menjadi sumber data berharga untuk penelitian dan analisis sentimen, membantu memahami persepsi dan pengalaman pengguna terhadap aplikasi tersebut. Melalui partisipasi pengguna dalam menyampaikan opini mereka, Gramedia Digital dapat memperbaiki dan mengembangkan produk dan layanan mereka sesuai dengan kebutuhan dan harapan pengguna.

*K-Nearest Neighbour* adalah sebuah model klasifikasi menggunakan konsep supervised learning untuk mengidentifikasi kategori atau label suatu objek berdasarkan variabel dan data latih yang ada. Proses klasifikasi ini melibatkan penggunaan data latih sebagai referensi untuk mengenali pola dalam data dan menghasilkan prediksi yang tepat. Untuk melakukan klasifikasi pada data baru (data uji), model menggunakan metode perhitungan jarak *Euclidean* untuk menemukan objek data latih yang paling mirip dengan data baru tersebut. Dengan demikian, model dapat mengklasifikasikan data baru berdasarkan kesamaannya dengan data latih yang telah diketahui labelnya [2]. Berdasarkan penelitian yang dilakukan sebelumnya oleh N.Habibah [3] salah satu keunggulan dari KNN adalah kemampuannya untuk menghasilkan prediksi yang lebih akurat dan efisien ketika memiliki jumlah data latih yang besar. Metode Lexicon Based adalah salah satu pendekatan yang digunakan untuk melakukan klasifikasi pada kalimat yang mengandung opini atau tidak, dengan mengacu pada kesesuaian kata-kata dalam dataset Lexicon berbahasa Indonesia. Setiap kata dalam Lexicon akan memiliki nilai polaritas yang menunjukkan sentimen atau perasaan yang terkandung dalam kalimat tersebut. Dengan demikian, metode ini memungkinkan untuk menentukan apakah kalimat tersebut bersifat positif, negatif, atau netral berdasarkan nilai polaritas kata-kata yang ada dalam kalimat tersebut [2]. Metode *Lexicon Based* digunakan sebagai alat pengklasifikasi sentimen dari setiap opini, memungkinkan kalimat sentimen untuk diklasifikasikan ke dalam kategori negatif dan positif. Dengan menggunakan metode ini, analisis sentimen pada data kalimat menjadi lebih sederhana dan mudah dilakukan [4]. Walaupun metode ini menunjukkan tingkat presisi yang tinggi dalam mengidentifikasi sentimen, namun nilai *recall*-nya cenderung rendah dikarenakan keterbatasan *Lexicon* yang digunakan sebagai pengenalan. Oleh karena itu, penulis melakukan kombinasi metode *Lexicon Based* dengan KNN untuk mengoptimalkan hasil identifikasi pada teks.

Penelitian sebelumnya tentang Analisis Sentimen yang menggunakan metode KNN dan *Lexicon Based* diantaranya adalah membahas tentang Analisis Sentimen Masyarakat Indonesia Terhadap Pembelajaran Online Dari Media Sosial Twitter Menggunakan *Lexicon* dan *K-Nearest Neighbour* [2], yang menyatakan bahwa melalui perhitungan metode *K-Nearest Neighbour* dengan *confusion matrix* untuk *tweet* OVO diperoleh nilai akurasi positif sebesar 86,91% dan nilai akurasi negatif 13,09%, *tweet* Gopay diperoleh nilai akurasi sebesar 94,05% dan nilai akurasi negatif 5,95% serta *tweet* Link Aja diperoleh nilai akurasi sebesar 76,31% dan nilai akurasi negatif 23,69%. Dalam penelitian lain berjudul Analisis Sentimen pada Aplikasi Peduli Lindungi dengan Menggunakan Metode *Improved K-Nearest Neighbour* dan *Lexicon Based* [5], menyatakan menggunakan metode *Improved K-Nearest Neighbour* memperoleh nilai akurasi tertinggi pada *k-values* 20 sebesar 85%. Pada dua penelitian tersebut dilakukan pelabelan sebelum tahapan preprocessing. Pada penelitian ini, terdapat perbedaan dalam metode pelabelan yang digunakan. Pada penelitian ini, pelabelan dilakukan setelah tahapan preprocessing selesai dilakukan. Perbedaan lain juga terdapat pada dataset kamus stopword yang digunakan. Pada penelitian ini, penulis menggunakan daftar kata stopword dari penelitian Fadillah Z. Tala [6].

Penelitian ini bertujuan untuk mendapatkan kesimpulan terkait dengan kepuasan pengguna terhadap aplikasi Gramedia Digital, serta mengetahui hasil akurasi nilai dari metode *K-Nearest Neighbour* dan *Lexicon Based Dataset* yang digunakan diambil dari teks *tweet* yang bersumber dari media sosial Twitter dengan menggunakan kata kunci ‘Gramedia Digital’, ‘Gramdig’ pada periode 01 Januari 2023 hingga 15 Mei 2023 dengan total 1330 *tweet*. Hasil analisis tersebut menghasilkan 2 sentimen yaitu Positif dan Negatif.

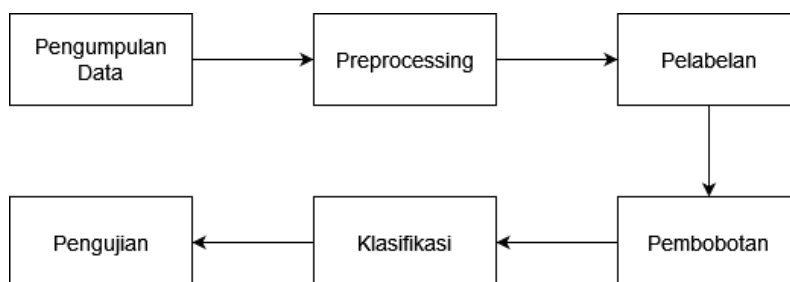
## 2. METODE PENELITIAN

### 2.1 Data Penelitian

Sumber data untuk penelitian ini adalah dari media sosial Twitter. Data yang digunakan berasal dari *tweets* yang dibuat oleh pengguna Aplikasi Gramedia Digital. Pengumpulan data dilakukan menggunakan pustaka *snsrape* dengan teknik *scraping*. Total terdapat 1330 *tweets* yang berhasil dikumpulkan dalam periode 01 Januari 2023 hingga 15 Mei 2023. Dalam proses pengumpulan data, digunakan 2 kata kunci yaitu ‘Gramedia Digital’ dan ‘Gramdig’. Selanjutnya, data tersebut melalui tahapan *preprocessing* agar dapat digunakan sebagai *dataset* dalam penelitian ini.

### 2.2 Penerapan Metode

Dalam membuat sistem analisis sentimen pada penelitian ini yang menggunakan metode *K-Nearest Neighbour*, beberapa langkah dilakukan untuk memastikan bahwa penelitian ini relevan dengan tujuan dan terukur. Tahap-tahap ini mewakili setiap desain dan proses dari awal hingga akhir sistem yang sedang berjalan.



Gambar 2.1 Tahapan Metode

Tahap pertama adalah pengumpulan data yang dilakukan menggunakan proses *scraping* dari media sosial Twitter untuk mendapatkan data. Kemudian hasil *scraping* yang tersimpan dengan format csv untuk diimpor ke dalam *database* untuk dilakukan tahap *preprocessing* melalui sistem analisis sentimen berbasis web.

*Preprocessing* adalah proses perubahan bentuk data yang terstruktur sembarang menjadi data yang terstruktur sesuai kebutuhan untuk proses dalam *text mining* [7]. Tujuan dari tahap *preprocessing* adalah untuk membersihkan dataset yang didapatkan dari media sosial Twitter, sehingga proses pelabelan dan pengujian hasil klasifikasi sentimen dapat dilakukan secara tepat dan akurat. Berikut adalah proses-proses yang digunakan penelitian ini pada tahap *preprocessing*:

a. *Case Folding*

*Case Folding* yaitu proses mengonversi huruf pada *dataset* menjadi huruf kecil (*lowercase*) untuk memperoleh keseragaman dalam teks [8]. Contoh hasil *case folding* misalnya ‘Gramedia’ menjadi ‘gramedia’, ‘NOVEL’ menjadi ‘novel’, dan seterusnya.

b. *Cleansing*

*Cleansing* yaitu proses membersihkan atau menghilangkan karakter yang tidak relevan atau tidak diperlukan dalam data *tweet*, seperti tanda baca, angka, URL, *username*, *mention*, *hashtag*, dan *retweet* seperti (“”~&?~><#%{}{[0-9]+;,:’) [9].

c. Normalisasi

Setelah proses *cleansing* selesai, langkah berikutnya dalam penelitian ini adalah normalisasi. Tujuannya adalah untuk mengubah setiap kata yang mengandung kata gaul, kata tidak baku, atau singkatan tertentu menjadi kata baku yang lebih standar. Normalisasi dilakukan dengan merujuk pada kamus *slang words* yang ada dalam *database*, serta penambahan manual kata-kata terkait dengan Aplikasi Gramedia Digital yang sedang diteliti. Sebagai contoh, kata “bca” diubah menjadi “baca,” kata “blomm” menjadi “belum,” dan seterusnya.

d. *Noise removal*

Setelah proses normalisasi selesai, penelitian ini melanjutkan dengan tahap penghapusan *stopwords* atau kata-kata yang tidak relevan. *Stopwords* merupakan kata-kata umum yang sering muncul dalam dokumen dan dianggap tidak memiliki kontribusi signifikan dalam analisis sentimen. Penelitian ini mengacu pada penjelasan pada sub-bab (2.4.4), di mana penghapusan *stopwords* menggunakan daftar *stopwords* berbahasa Indonesia yang diperoleh dari penelitian Fadillah Z. Tala [6]. Daftar ini juga diperbarui secara manual dengan menambahkan kata-kata yang relevan dengan konteks Aplikasi Gramedia Digital yang sedang diteliti. Sebagai contoh, jika kata “adalah” ada dalam daftar *stopwords*, maka kata tersebut akan dihapus dari dokumen.

e. *Stemming*

*Stemming* adalah proses mengubah kata ke bentuk dasarnya, yang bertujuan untuk mengurangi variasi kata menjadi kata dasar yang sama. Sebagai contoh, pada kata "membaca", setelah melalui proses *stemming*, kata tersebut akan menjadi "baca" [10]. Untuk melakukan proses ini, digunakan pustaka Sastrawi yang mengimplementasikan Algoritma Nazief dan Adriani.

f. *Tokenizing*

Tahap terakhir adalah *tokenizing* yang melakukan proses memecah teks menjadi unit-unit yang lebih kecil yang disebut token. Misalnya, dalam kalimat “Saya baca di gramedia digital”, tokenisasi akan memisahkan kata-kata tersebut menjadi token individu: “Saya”, “baca”, “di”, “gramedia”, “digital”. Dengan memecah teks menjadi token-token terpisah, dapat dilakukan analisis lebih lanjut seperti menghitung frekuensi kata.

Pelabelan sentimen pada penelitian ini dilakukan secara otomatis, yaitu melibatkan perhitungan nilai skor dengan memanfaatkan kamus *lexicon* yang menggunakan dataset <https://github.com/fajri91/InSet.git>. Dalam metode perhitungan skor pelabelan sentimen, dalam pendekatannya digunakan metode untuk menghitung skor nilai sentimen kelas positif dan menguranginya dengan skor nilai sentimen kelas negatif pada setiap tweet. Proses pelabelan dilakukan secara otomatis dengan menggunakan bahasa pemrograman Python, di mana jika skor suatu kalimat > 0, maka akan dikategorikan sebagai positif. Jika skor kalimat tersebut < 0, maka akan dikategorikan sebagai negatif. Sedangkan jika skornya sama dengan 0, maka akan dikategorikan sebagai netral.

Pada tahap *K-Nearest Neighbour*, metode klasifikasi digunakan untuk melakukan analisis sentimen. Tahap ini melibatkan pelatihan dokumen yang telah melalui tahap pelabelan, *preprocessing*, dan proses perhitungan bobot kata dengan menggunakan pembobotan TF-IDF dari data latih dan data uji. Adapun klasifikasi *K-Nearest Neighbour* dapat dilihat pada gambar 2.2:



Gambar 2. 2 Tahapan *K-Nearest Neighbour*

Proses dilanjutkan dengan menghitung jarak antara data terbaru dengan setiap label data menggunakan seluruh model data latih. Kemudian, untuk menentukan tingkat kesamaan pada dokumen tersebut, digunakan metode *euclidean distance* sesuai dengan rumus berikut:

$$D(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 \dots + (x_k - y_k)^2} = \sqrt{\sum_{k=1}^n (X_k - Y_k)^2} \quad (1)$$

Pengujian pada penelitian ini menggunakan confusion matrix untuk membandingkan data dari hasil prediksi dengan sekumpulan data aktual. Tujuan dari pengujian ini adalah untuk memperoleh nilai atau tingkat akurasi, presisi, dan *recall* dari klasifikasi yang menggunakan metode *K-Nearest Neighbour*. Penelitian ini fokus pada dua kelas sentimen, yaitu positif dan negatif.

a. Akurasi

Akurasi adalah ukuran yang mencerminkan jumlah dokumen yang diklasifikasikan dengan benar, termasuk *True Positive* dan *True Negative*. Untuk menghitung nilai akurasi, digunakan persamaan berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN} \times 100 \% \quad (2)$$

1. *True Positive*

*True Positive* adalah data positif yang telah berhasil diprediksi dengan benar oleh model. Contohnya, jika sebuah *tweet* memiliki label positif dan model yang telah dibuat berhasil memprediksi *tweet* tersebut sebagai positif.

2. *True Negative*

*True Negative* merujuk pada data negatif yang berhasil diprediksi dengan benar. Sebagai contoh, jika sebuah *tweet* memiliki label negatif dan model yang telah dibuat berhasil memprediksi *tweet* tersebut sebagai negatif.

3. *False Positive*

*False Positive* adalah data negatif yang keliru diprediksi sebagai positif oleh model. Sebagai contoh, ketika sebuah *tweet* memiliki label negatif, namun model yang dibuat dengan salah memprediksi *tweet* tersebut sebagai positif.

4. *False Negative*

*False Negative* adalah data positif yang keliru diprediksi sebagai negatif oleh model. Sebagai contoh, ketika sebuah *tweet* memiliki label positif, namun model yang dibuat dengan salah memprediksi *tweet* tersebut sebagai negatif.

b. Presisi

Presisi adalah berapa banyak hasil pemrosesan relevan dengan informasi yang dicari. Dapat juga diartikan presisi merupakan klasifikasi *True Positive* dan semua data diprediksi sebagai kelas positif. Untuk mendapatkan nilai presisi dapat menggunakan persamaan berikut ini:

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100 \% \quad (3)$$

c. *Recall*

*Recall* merupakan persentase dokumen yang secara benar diklasifikasikan sebagai *True Positive* dari total jumlah dokumen yang benar-benar positif (termasuk yang diklasifikasikan secara *False Negative*). Untuk mendapatkan nilai *recall* dapat menggunakan persamaan berikut ini:

$$\text{Recall} = \frac{TP}{TP+FN} \times 100 \% \quad (4)$$

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Implementasi Metode

Tahap ini merupakan langkah awal dalam menentukan *dataset* yang akan digunakan. *Dataset* yang digunakan diperoleh menggunakan pustaka *snsrape* melalui proses *scraping* dari periode 01 Januari 2023 sampai dengan 15 Mei 2023 dengan kata kunci yaitu 'Gramedia Digital' dan 'Gramdig'. Setelah data *tweet* berhasil didapatkan dari proses *scraping* lalu disimpan dalam file berformat CSV, *dataset* akan diimpor ke dalam *database* untuk dilakukan

proses lanjutan. Contoh data sampel *tweet* pengguna Aplikasi Gramedia Digital dapat dilihat pada tabel berikut ini:

**Tabel 3. 1 Data Sampel *Tweet* Pengguna Aplikasi Gramedia Digital**

No.	Created at	Username	Tweets
1	30/04/2023 20:54	irkustk	Iseng beli paket promo gramedia digital yg buku2 fiksi, isinya Harlequin banyak bener 😊
2	30/04/2023 21:41	purplefluffy_	Pasta Kadang Merah • Durian Sukegawa • 240 halaman • Penerbit: @bukugpu • Penerjemah: Asri Pratiwi Wulandari   Genre: Contemporary Fiction, Slice of Life, j-lit   Baca di @GramediaDigital <a href="https://t.co/7aN8ZmjFR">https://t.co/7aN8ZmjFR</a>
No.	Created at	Username	Tweets
3	30/04/2023 19:20	moonstresss	@gramediadotcom @GramediaDigital Aku masih setia menunggu kehadiran mereka inih 🙏🙏🙏 <a href="https://t.co/69aBtRoR84">https://t.co/69aBtRoR84</a>
...	...	...	...
1330	01/01/2023 07:13	queerindisguise	@cm_bwtsng Halo, maaf baru dibalas, ini seri #YaTuhan ya, ada di Gramedia Digital, yang di foto itu seri buku ponsel, .3gp, doa. Semoga membantu 😊 <a href="https://t.co/ORZqq4PyBq">https://t.co/ORZqq4PyBq</a>

### 3.2 Preprocessing

Proses tahapan preprocessing data kepuasan pengguna aplikasi Gramedia Digital tersebut dapat dilihat pada tabel berikut ini:

**Tabel 3. 2 Preprocessing *Tweet* Pengguna Aplikasi Gramedia Digital**

Tahap	Teks Awal	Teks Proses
<i>Case folding</i>	Tampilan Gramedia Digital yang sekarang ini emang jelek pisan, udahlah makin mihil *eh Berharap bakal balik ke layout ebook sebelumnya, tapi kok tydac balik <sup>2</sup> ~ 😊 <a href="https://t.co/QQ1ShxzZFP">https://t.co/QQ1ShxzZFP</a>	tampilan gramedia digital yang sekarang ini emang jelek pisan, udahlah makin mihil *eh berharap bakal balik ke layout ebook sebelumnya, tapi kok tydac balik <sup>2</sup> ~ 😊 <a href="https://t.co/qq1shxzZFP">https://t.co/qq1shxzZFP</a>
<i>Cleansing</i>	tampilan gramedia digital yang sekarang ini emang jelek pisan, udahlah makin mihil *eh berharap bakal balik ke layout ebook sebelumnya, tapi kok tydac balik <sup>2</sup> ~ 😊 <a href="https://t.co/qq1shxzZFP">https://t.co/qq1shxzZFP</a>	tampilan gramedia digital yang sekarang ini emang jelek pisan udahlah makin mihil eh berharap bakal balik ke layout ebook sebelumnya tapi kok tydac balik
Normalisasi	tampilan gramedia digital yang sekarang ini emang jelek pisan udahlah makin mihil eh berharap bakal balik ke layout ebook sebelumnya tapi kok tydac balik	tampilan gramedia digital yang sekarang ini memang jelek sekalian sudah makin mahal eh berharap bakal balik ke tampilan buku elektronik sebelumnya tapi kok tidak balik
Menghapus Kata <i>Stop</i>	tampilan gramedia digital yang sekarang ini memang jelek sekalian sudah makin mahal eh berharap bakal balik ke tampilan buku elektronik sebelumnya tapi kok tidak balik	tampilan gramedia digital jelek mahal berharap tampilan buku elektronik
<i>Stemming</i>	tampilan gramedia digital jelek mahal berharap tampilan buku elektronik	tampil gramedia digital jelek mahal harap tampil buku elektronik

*Tokenizing* tampil gramedia digital jelek mahal harap tampil, gramedia, digital, jelek, mahal, harap, tampil buku elektronik tampil, buku, elektronik

### 3.3 Pengujian Algoritma

Pada tahap pengujian, merupakan bagian penting yang harus dilakukan dalam setiap pengembangan sistem untuk mengevaluasi tingkat keberhasilan dari implementasi dengan menggunakan kriteria atau metrik yang telah ditentukan sebelumnya. Penelitian ini melakukan pengujian terhadap kepuasan pengguna aplikasi Gramedia Digital berdasarkan perhitungan akurasi, presisi, dan juga *recall*. Seperti yang sudah dijelaskan di batasan masalah, nilai  $k$  yang sudah ditentukan adalah  $k = 3$ ,  $k = 5$ ,  $k = 7$ ,  $k = 9$ , sehingga akan dilakukan perbandingan dengan rasio pembagian *dataset* sehingga mendapatkan nilai akurasi, presisi, dan *recall* yang terbaik.

**Tabel 3. 3 Hasil Pengujian Pembagian Data Kepuasan Pengguna Aplikasi Gramedia Digital**

	Rasio	09:01	08:02	07:03	06:04	05:05
Akurasi	K = 3	70,3%	71,76%	74,30%	75,97%	74,13%
	K = 5	67,18%	70,57%	71%	72,50%	70,83%
	K = 7	68,73%	67%	69,50%	71%	70,50%
	K = 9	64,86%	67,32%	69,85%	73,66%	69,45%

Dari data yang tercantum dalam Tabel 3.3, dapat disimpulkan bahwa penggunaan rasio 60% data latih dan 40% data uji menghasilkan akurasi tertinggi, mencapai 75,97%. Hasil prediksi juga terdokumentasi dalam tabel berikut.

**Tabel 3. 4 Data Sampel Hasil Prediksi Kepuasan Pengguna Aplikasi Gramedia Digital**

No.	<i>Tweets</i>	Label (Aktual)	Label (Prediksi)
1.	lihat <i>website</i> gramedia digital cari ebooknya terang masuk penuh premium premium fiksi beli satu	<i>positive</i>	<i>positive</i>
2.	malam begadang jam pagi lanjut pagi selesai not folks ayu welirang baca gramedia digital	<i>negative</i>	<i>positive</i>
3.	gramedia digital jarang jajan buku distraksinya tambah aplikasi tonton timbun gabung cetak digital keluar rumah presiden periode	<i>negative</i>	<i>positive</i>
4.	tampil gramedia digital jelek udahlah mahal harap tampil buku elektronik	<i>negative</i>	<i>negative</i>
5.	gramedia digital rusak	<i>negative</i>	<i>negative</i>
...	...	...	...
225	sewa gramedia digital minus baca elektronik pakai buku	<i>negative</i>	<i>positive</i>

Berdasarkan tabel 3.4 di atas, pengujian tersebut menggunakan data uji dengan nilai 225 data. Selanjutnya hasil perhitungan pengujian dipresentasikan dalam bentuk tabel *confusion matrix*. Tabel *confusion matrix* kepuasan pengguna aplikasi Gramedia Digital dapat dilihat pada tabel berikut ini.

**Tabel 3.5 Confusion Matrix Kepuasan Pengguna Aplikasi Gramedia Digital**

Data Aktual	Data Prediksi		Total
	Positif	Negatif	
Positif	111	121	232
Negatif	45	414	459
Total	156	535	691

Berdasarkan dari tabel 3.5 di atas, maka nilai akurasi, presisi, dan *recall* sudah dijelaskan dalam persamaan rumus (2), persamaan rumus (3), dan persamaan rumus (4). Hasil perhitungan pengujian kepuasan pengguna aplikasi gramedia digital dapat dilihat pada tabel berikut ini.

**Tabel 3.6 Tabel Hasil Perhitungan Kepuasan Pengguna Aplikasi Gramedia Digital**

Akurasi	Presisi	Recall
$= ((TP + TN) / (TP + TN + FP + FN)) \times 100\%$	$= (((TP / (TP + FP)) + (TN / (TN + FN))) / 2) \times 100\%$	$= (((TP / (TP + FN)) + (TN / (TN + FP))) / 2) \times 100\%$
$= ((111 + 414) / (111 + 414 + 45 + 121)) \times 100\%$	$= (((111 / (111 + 45)) + (121 / (121 + 414))) / 2) \times 100\%$	$= (((111 / (111 + 414)) + (121 / (121 + 45))) / 2) \times 100\%$
75,97%	$= ((0.71) + (0.74) / 2) \times 100\%$ 77,38%	$= ((0.73) + (0.72) / 2) \times 100\%$ 90,19%

Berdasarkan tabel 3.6, sudah diketahui hasil pengujian kepuasan pengguna aplikasi Gramedia Digital menggunakan Algoritma *K-Nearest Neighbour* mendapatkan nilai akurasi 75,97%, presisi 77,38%, dan juga *recall* 90,19%.

#### 4. KESIMPULAN

Setelah melakukan penelitian dan analisis terhadap sistem Analisis Sentimen terhadap kepuasan pengguna aplikasi Gramedia Digital yang telah dikembangkan, dapat disimpulkan bahwa terdapat beberapa kesimpulan dan saran untuk meningkatkan sistem Analisis Sentimen yaitu: Dalam *dataset* yang terdiri dari 1330 *tweets* dalam periode 01 Januari 2023 hingga 15 Mei 2023, metode klasifikasi berhasil mencapai akurasi tertinggi sebesar 75,97% dengan menggunakan nilai  $k = 3$ , serta membagi data latih dan data uji dengan rasio 60% dan 40% secara berturut-turut. Pada penelitian ini, pelabelan menggunakan metode Lexicon Based dan klasifikasi menggunakan metode *K-Nearest Neighbours* menghasilkan akurasi yang cukup baik, tetapi dikarenakan banyaknya dataset yang menggunakan campuran bahasa Indonesia dan bahasa Inggris, penerapan proses *case folding*, *cleansing*, normalisasi, penghapusan kata stop, *stemming*, dan *tokenizing* memberikan dampak yang signifikan dalam meningkatkan akurasi klasifikasi. Sistem Analisis Sentimen ini telah dikembangkan sebagai sebuah alat yang berguna dalam memberikan informasi kepada pihak terkait. Tujuannya adalah untuk menyediakan data yang dapat digunakan sebagai dasar pertimbangan dalam menganalisis sentimen yang berasal dari masyarakat.

#### DAFTAR PUSTAKA

- [1] I. Utami and M. Marzuki, "Analisis Sistem Informasi Banjir Berbasis Media Twitter," *Jurnal Fisika Unand*, vol. 9, no. 1, pp. 67–72, 2020, doi: 10.25077/jfu.9.1.67-72.2020.
- [2] A. Halimi, K. Kusri, and M. R. Arief, "Analisis Sentimen Masyarakat Indonesia Terhadap Pembelajaran Online Dari Di Media Sosial Twitter Menggunakan Lexicon Dan K-Nearest Neighbor," *COREAI: Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi*, vol. 2, no. 1, pp. 18–28, 2021, doi: 10.33650/coreai.v2i1.2283.



- [3] N. Habibah, E. Budianita, M. Fikry, and I. Iskandar, “Analisis Sentimen Mengenai Penggunaan E-Wallet Pada Google Play Menggunakan Lexicon Based dan K-Nearest Neighbor,” vol. 10, no. 1, pp. 192–200, 2023, doi: 10.30865/jurikom.v10i1.5429.
- [4] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan, “Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine,” *Komputek*, vol. 3, no. 2, p. 52, 2019, doi: 10.24269/jkt.v3i2.270.
- [5] N. Putriani, F. R. Umbara, and P. N. Sabrina, “Analisis Sentimen pada Aplikasi PeduliLindungi dengan Menggunakan Metode Improved K-Nearest Neighbor dan Lexicon Based,” vol. 8, no. 1, pp. 350–364, 2022.
- [6] F. Z. Tala, “A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia,” *M.Sc. Thesis, Appendix D*, vol. pp, pp. 39–46, 2003.
- [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: Jurnal Pendidikan Informatika*, vol. 6, no. 1, pp. 98–106, 2022, doi: 10.29408/edumatic.v6i1.5433.
- [8] D. A. Lestari and D. Mahdiana, “Penerapan Algoritma K-Nearest Neighbor pada Twitter untuk Analisis Sentimen Masyarakat Terhadap Larangan Mudik 2021,” *Informatik : Jurnal Ilmu Komputer*, vol. 17, no. 2, p. 123, 2021, doi: 10.52958/iftk.v17i2.3629.
- [9] M. Furqan, S. Sriani, and S. M. Sari, “Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia,” *Techno.Com*, vol. 21, no. 1, pp. 51–60, 2022, doi: 10.33633/tc.v21i1.5446.
- [10] F. Rizqi Irawan, “Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearset Neighbors,” *JIKO (Jurnal Informatika dan Komputer)*, vol. 5, no. 1, pp. 62–68, 2022, doi: 10.33387/jiko.v5i1.4267.