



Analisis Sentimen Pada Komentar Instagram *Provider* By.U Menggunakan Metode *K-Nearest Neighbors* (KNN)

Nindy Alvionika^{1*}, Sutan Faisal², Rahmat³, Anis Fitri Nur Masruriyah⁴

^{1,2,3,4}Universitas Buana Perjuangan Karawang

**email*: if20.nindyalvionika@mhs.ubpkarawang.ac.id

Info Artikel	ABSTRAK
Dikirim: 3 September 2023 Diterima: 14 April 2024 Diterbitkan: 31 Mei 2024	Kehadiran internet dari berbagai penyedia layanan menjadi semakin beragam, menuntut kualitas unggul untuk memenuhi kebutuhan pengguna. Salah satu penyedia layanan internet terkemuka dan terbesar di Indonesia yaitu Telkomsel, yang kini memiliki <i>sub-brand</i> bernama By.U. By.U resmi diperkenalkan pada 10 Oktober 2019 dan merupakan salah satu ISP (<i>Internet Service Provider</i>) pertama di Indonesia yang menawarkan pengalaman digital secara (E2E) <i>End-To-End</i> pada seluruh aspek kebutuhan telekomunikasi. Pengalaman digital (E2E) pada By.U memudahkan pengguna untuk memilih metode pengiriman kartu perdana hingga menentukan nomor telepon By.U, membeli paket data, dan menyelesaikan proses pembayaran. Menganalisis sentimen dapat mengidentifikasi representasi opini pengguna berdasarkan pengalaman pengguna By.U. Metode <i>K-Nearest Neighbors</i> yang akan digunakan dalam penelitian ini. Berdasarkan penelitian yang telah dilakukan, hasil akurasi menggunakan algoritma <i>K-Nearest Neighbors</i> sebesar 73 persen. Hasil pada <i>topic modelling</i> yang dilakukan menggunakan <i>Latent Dirichlet Allocation</i> diketahui tema yang dibahas pada topik 1 yaitu menggambarkan mengenai <i>information</i> data telekomunikasi, pada topik 2 mengenai pendaftaran sim <i>card</i> , topik 3 mencakup berbagai masalah yang berkaitan dengan paket internet, aplikasi, dan transaksi. Hasil analisis sentimen menunjukkan bahwa penelitian ini mengalami kesulitan dalam memprediksi sentimen, tidak ada prediksi yang akurat untuk kelas positif, netral dan negatif. Hal ini menunjukkan bahwa penelitian harus ditingkatkan dan dilatih dengan menggunakan sampel yang lebih besar untuk prediksi sentimen.
Kata kunci: Aplikasi; By.U; Internet; <i>K-Nearest Neighbors</i> ; Telkomsel.	

1. PENDAHULUAN

Internet telah tumbuh menjadi bagian integral dari kehidupan sehari-hari masyarakat di era revolusi industri keempat [1] yang menjadi sarana penting dalam berkomunikasi, bekerja, belajar, dan berbelanja. Kehadiran internet dari berbagai penyedia layanan menjadi semakin beragam, menuntut kualitas unggul untuk memenuhi kebutuhan pengguna. Salah satu ISP (*Internet Service Provider*) yaitu Telkomsel, yang kini memiliki *sub-brand* bernama By.U [2]. By.U resmi diperkenalkan pada 10 Oktober 2019 dan merupakan salah satu ISP pertama di Indonesia yang menawarkan pengalaman digital secara (E2E) *End-To-End* pada seluruh aspek kebutuhan telekomunikasi [3]. Pengalaman digital (E2E) pada By.U memudahkan pengguna untuk memilih metode pengiriman kartu perdana hingga menentukan nomor telepon By.U, membeli paket data, dan menyelesaikan proses pembayaran. Kuota internet By.U dapat dinikmati pengguna sepanjang waktu, 24 jam penuh, pada semua sinyal Telkomsel (2G, 3G, 4G) [4]. Selama empat tahun terakhir, beragam fasilitas yang ditawarkan oleh penyedia By.U telah menarik banyak pengguna baru yang sebagian besar adalah Gen Z.

Provider By.U mengutamakan fokusnya terhadap Gen Z yang semakin sadar dan paham tentang dunia digital. By.U sebagai perusahaan komunikasi digital menawarkan beragam keunggulan dan pengalaman *seamless* yang memberikan kemudahan bagi pengguna [5]. Seluruh aktivitas layanan dilakukan dalam aplikasi By.U yang terinstall pada *smartphone* yang dapat diunduh di Google *Play Store* dan *App Store*. Hal ini memudahkan pengguna untuk mengontrol layanan seluler yang diinginkan sesuai dengan kebutuhannya. Meskipun By.U dilabeli yang serba digital dengan menawarkan kemudahan, tetapi tidak semua pengguna merasa puas atau terlayani dengan baik. Adapun keterbatasan yang dimiliki By.U seperti respon Nindy sebagai admin mungkin tidak dapat memahami pertanyaan pengguna. Pengguna harus menjelaskan masalah atau keluhannya secara detail. Jika pertanyaan tidak terjawab, pengguna dapat menghubungi *call center*. Namun hanya pengguna dengan nomor By.U yang dapat menghubungi *call center*. Layanan pengguna By.U dapat digunakan kapanpun dan dimanapun apabila membutuhkan layanan By.U yang tersedia 24 jam sehari melalui *live chat* dan *call center*. Dibutuhkan *developer* untuk menjaga kualitas dan perkembangan penyedia layanan aplikasi *provider* By.U.

Semakin banyaknya pengguna internet yang beralih ke layanan digital By.U, sebanyak 20 juta pengguna aplikasi *provider* By.U [6] yang akhirnya membagikan komentar melalui media sosial Instagram. Instagram merupakan *platform* media sosial yang banyak digunakan oleh berbagai perusahaan untuk mempromosikan *provider* By.U dan berinteraksi langsung dengan *costumer* [7]. By.U menggunakan Instagram untuk menjangkau audiens yang lebih besar dan mengumpulkan umpan balik *costumer*. Pengguna internet yang beralih ke layanan digital By.U meninggalkan komentar di postingan Instagram By.U dengan berbagai komentar, mulai dari pujian hingga keluhan. Dalam proses NLP (*Natural Language Processing*), analisis sentimen merupakan metode teknik yang digunakan untuk menganalisis dan mendeskripsikan pendapat atau ekspresi manusia dalam dokumen [8]. Analisis sentimen dapat menampilkan suatu objek atau topik dalam format teks maupun dokumen. Sentimen ini dapat mengidentifikasi pendapat seseorang dalam teks untuk mengevaluasi pendapat tersebut, yang dapat diklasifikasikan menjadi tiga kategori: positif, negatif, dan netral [9]. Hasil analisis sentimen dapat mengidentifikasi persepsi pengguna berdasarkan pengalaman By.U. Hasil analisis sentimen dapat memberikan gambaran mengenai pendapat pengguna berdasarkan pengalaman saat menggunakan aplikasi By.U. Metode *K-Nearest Neighbors* yang akan digunakan dalam penelitian ini. Pendekatan serupa yaitu menganalisis sentimen masyarakat terhadap LGBT di Indonesia menggunakan algoritma *K-Nearest Neighbors* menghasilkan akurasi yang cukup tinggi yaitu sebesar 85.90 persen [10].

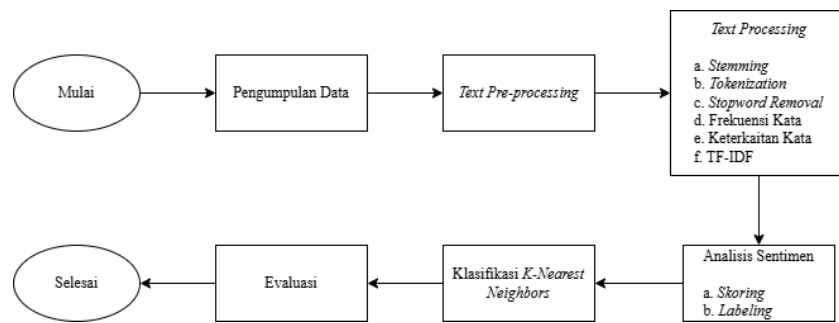
Penelitian dengan objek serupa dalam menganalisis sentimen pengguna By.U pada aplikasi Twitter menggunakan metode *Naïve Bayes*, dengan melakukan 3 kali pengujian menggunakan 3000 dataset yang menghasilkan nilai akurasi sebesar 85 persen [2]. Berdasarkan penelitian sebelumnya yang menganalisis sentimen masyarakat terhadap relokasi Ibukota Nusantara dengan membandingkan *Naïve Bayes* dan *K-Nearest Neighbors* [11]. Menampilkan bahwa hasil akurasi KNN lebih tinggi yaitu sebesar 88.12 persen. Kemudian, penelitian keempat yang melakukan perbandingan algoritma untuk menganalisis sentimen terhadap data ulasan aplikasi pinjaman online berizin OJK (Otoritas Jasa Keuangan). Menunjukkan bahwa hasil akurasi pada KNN lebih tinggi yaitu 93.75 persen, dibandingkan hasil akurasi Decision Tree yaitu 76.56 persen [12]. Penelitian selanjutnya menganalisis sentimen pengguna dalam membaca *review game* steam dengan menggunakan *Support Vector Machine* dan *K-Nearest Neighbors* serta memanfaatkan *text summarization* [13]. Hasil pengujian analisis sentimen menggunakan *k-cross validation* 5 menghasilkan nilai akurasi sebesar 62 persen, sedangkan analisis sentimen dan *text summarization* menghasilkan nilai akurasi sebesar 60 persen. Namun, penggunaan *text summarization* untuk melakukan *review game* secara keseluruhan tidak efektif dalam meningkatkan hasil analisis sentimen. Sehingga penelitian ini akan menggunakan metode *K-Nearest Neighbors* untuk mengklasifikasikan sentimen pada komentar Instagram By.U.

Oleh karena itu, penelitian ini berfokus pada sentimen aplikasi *provider* By.U berdasarkan komentar pengguna Instagram. Metode *K-Nearest Neighbors* digunakan untuk menganalisis data. Setelah mendapat hasil analisis sentimen, kemudian dilakukan *Topic Modelling* menentukan tema apa yang dibahas dalam komentar tersebut. *Topic Modelling* adalah model statistik yang digunakan untuk mengidentifikasi inti atau topik dalam suatu kumpulan dokumen [14]. Penelitian ini juga menguji seberapa akurat model dalam memprediksi dan mengklasifikasikan teks ke dalam sentimen positif dan negatif pada dataset By.U. Hasil penggunaan metode

Naïve Bayes Classifier pada klasifikasi ulasan By.U memperoleh akurasi sebesar 95 persen, sedangkan klasifikasi *K-Nearest Neighbor* memperoleh nilai akurasi sebesar 89% [15]. Maka, dalam penelitian ini akan menguji model dengan mengklasifikasikan teks ke dalam tiga kategori yaitu positif, negatif, dan netral.

2. METODE PENELITIAN

Dalam penelitian ini memiliki beberapa tahap yang dilalui pada Gambar 1. Tahap-tahap tersebut meliputi pengumpulan data, *text pre-processing*, *text processing*, menganalisis sentimen, pemodelan *K-Nearest Neighbors* (KNN) dan evaluasi.



Gambar 1. Tahapan dalam penelitian

2.1 Pengumpulan Data

Proses pengumpulan dataset menggunakan teknik *crawling* data pada komentar foto/video di Instagram. Instagram dipilih karena popularitasnya di Indonesia [16], yang menempati posisi keempat di dunia dengan jumlah sebanyak 85.19 juta pada Januari 2023 [17]. Data komentar diperoleh dari akun @byu.id dengan foto atau video yang terkait dengan *provider* By.U supaya relevan dengan penelitian ini.

2.2 Text Pre-processing

Text pre-processing merupakan tahap krusial dalam mengorganisir data teks dan mengklasifikasikan informasi. Tahapnya terdiri dari *cleaning* (menghapus data tidak relevan) dan *case folding* (penyeragaman huruf). Ini membantu pengguna untuk mengakses informasi dengan lebih mudah.

2.3 Text Processing

Text processing merupakan proses mempersiapkan dokumen yang terdiri dari teks yang sebelumnya tidak terorganisir dan diubah menjadi teks yang lebih terorganisir sehingga dapat diproses [18]. Berikut ini merupakan proses yang dilakukan *text processing*, antara lain sebagai berikut:

- 1) *Stemming*
Pada langkah ini, setiap kata dalam kalimat diubah ke dalam bentuk dasar atau menghapus kata imbuhan.
- 2) *Tokenization*
Pada langkah ini, kalimat dipisahkan menjadi potongan kata atau token, dan selanjutnya kata yang tidak penting seperti nama pengguna dan angka akan dihilangkan.
- 3) *Stopword Removal*
Dalam langkah ini, kata-kata dalam dokumen teks yang tidak umum dan tidak memiliki makna akan dihapus.
- 4) *Frekuensi Kata*
Pada langkah ini, menghitung berapa jumlah setiap kata yang muncul dalam teks dan mengidentifikasi kata-kata yang biasa digunakan mungkin merupakan kata kunci penting.
- 5) *Keterkaitan Kata*

Pada langkah ini, melakukan identifikasi kata-kata yang sering muncul bersamaan dan membuat *wordcloud* untuk membantu memahami hubungan yang ada antara kata-kata tersebut dan topik yang dibahas dalam komentar.

6) TF-IDF (*Term Frequency – Inverse Document Frequency*)

Tahap berikutnya yaitu proses pembobotan menggunakan TF-IDF yaitu menghitung frekuensi term (tf) dalam dokumen (d), diikuti dengan df, idf, dan tf-idf. *Term frequency* (tf) adalah frekuensi kemunculan *term* (t) sebagai banyaknya kata dalam dokumen (d). Bobot ini digunakan pada algoritma K-Nearest Neighbors (KNN). TF-IDF dapat dirumuskan sebagai berikut [19]:

$$Tf.Idf = tf \times idf \quad (1)$$

Dalam pendekatan ini, menganggap dokumen sebagai vektor yang berisi komponen yang sesuai dengan setiap istilah dalam kamus dan bobot untuk setiap komponen. Istilah *term* yang tidak ditemukan dalam dokumen mempunyai bobot 0 [19].

2.4 Analisis Sentimen

Dalam penelitian ini, analisis sentimen bertujuan untuk membagi komentar pengguna Instagram terhadap provider By.U menjadi tiga kategori, yaitu positif, negatif atau netral. Untuk mencapai tujuan ini, analisis sentimen dibagi menjadi dua tahap yaitu *skoring* dan *labeling*.

1) *Skoring*

Dalam penelitian ini, *skoring* mengacu pada nilai numerik yang diberikan kepada setiap komentar untuk menunjukkan tingkat sentimen yang dimilikinya. Nilai ini berkisar dari nol hingga satu yang menunjukkan sentimen yang netral.

2) *Labeling*

Dalam penelitian ini, *labeling* mengacu pada proses memberi label sentimen (positif, negatif atau netral) kepada setiap komentar. *Labeling* dapat dilakukan secara otomatis menggunakan kamus positif dan negatif.

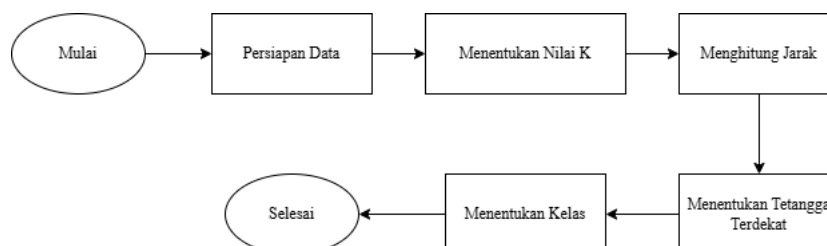
2.5 K-Nearest Neighbors

Metode K-Nearest Neighbor (K-NN) berdasarkan pada pembelajaran data yang sudah terklasifikasi untuk mengklasifikasikan sekumpulan data [20]. Metode *K-Nearest Neighbors* (KNN) adalah sebuah teknik pembelajaran diawasi yang mengklasifikasikan data menggunakan data terklasifikasi dengan jarak terdekat [21], dengan prinsip Euclidean Distance berdasarkan nilai k [22]. Metode *Euclidean Distance* digunakan untuk menghitung jarak antara dua titik dalam ruang multidimensional [23] yang dirumuskan sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=0}^n (x_{1i} - y_{1i})^2 + (x_{2i} - y_{2i})^2 + \dots} \quad (2)$$

Dimana $d(x, y)$ adalah mengukur jarak antara x dan y ; x adalah data yang akan diklasifikasi; y adalah data yang berada di sekitarnya; i adalah jumlah fitur dalam data.

K-Nearest Neighbors merupakan pendekatan untuk menemukan kasus baru dan lama berdasarkan perhitungan bobot yang sesuai. Algoritma KNN melibatkan beberapa tahap yaitu menentukan nilai K, menentukan nilai K maksimum, mengurutkan nilai K dari yang tertinggi hingga terendah, dan memprediksi kategori objek menggunakan kategori mayoritas tetangga terdekat [21]. Gambar 2. merupakan langkah-langkah metode K-Nearest Neighbor.



Gambar 2. Tahapan metode *k-nearest neighbors*

1) Persiapan Data

Pada proses persiapan data ini, telah melakukan *crawling* data pada komentar Instagram sebagai dataset yang akan dianalisis dalam penelitian ini.

2) Menentukan Nilai K

Mencari jumlah tetangga terdekat untuk membuat keputusan klasifikasi dan menemukan nilai K yang optimal menggunakan *cross-validation*. Nilai K biasanya digunakan sebagai bilangan ganjil untuk menghindari hasil klasifikasi seri.

3) Menghitung Jarak

Menghitung jarak antara titik data uji dan pelatihan menggunakan *Euclidean* untuk melakukan analisis kalimat baru dengan setiap kalimat dalam dataset pelatihan.

4) Menentukan Tetangga Terdekat

Berdasarkan jarak yang dihitung pada langkah sebelumnya, pilih k pada kalimat terdekat yang akan digunakan untuk menentukan sentiment kalimat baru.

5) Menentukan Kelas

Menentukan kelas yang paling sering muncul (*voting* mayoritas) dari K tetangga terdekat digunakan untuk mengklasifikasikan data baru.

2.6 Evaluasi

Penggunaan *confusion matrix* menghasilkan informasi nilai akurasi, presisi, dan *recall*. Semakin besar nilai AUC, maka semakin baik kemampuan prediktif model tersebut. Tabel 1. terdapat tabel *confusion matrix*.

Tabel 1. Tabel *confusion matrix*

Sumber: [24]

Aktual	Prediksi	
	True	False
True	TP	FN
False	FP	TN

Persamaan (3) hingga (5) menjelaskan perhitungan dalam *confusion matrix*. TP (*True Positive*) adalah prediksi positif yang tepat dari kelas positif, FP (*False Positive*) adalah prediksi positif yang sebenarnya dari kelas negatif. TN (*True Neagtive*) adalah prediksi negatif yang tepat dari kelas negatif, dan FN (*False Negative*) adalah prediksi negatif yang sebenarnya dari kelas positif [24].

$$Accuracy (A) = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (3)$$

$$Precision (P) = \frac{(TP)}{(TP + FP)} \quad (4)$$

$$Recall (R) = \frac{(TP)}{(TP + FN)} \quad (5)$$

Penggunaan informasi dalam *confusion matrix* dapat menghitung akurasi, *precision*, dan *recall*. Akurasi adalah seberapa tepatnya model dalam melakukan klasifikasi, dihitung dari jumlah prediksi benar dibagi dengan jumlah sampel. *Precision* mencerminkan ketepatan antara data yang diminta dan hasil prediksi model. *Recall* adalah rasio observasi positif yang diprediksi benar terhadap seluruh observasi dalam kelas sebenarnya [25].

Setelah menganalisis sentimen, dilakukan pemodelan topik dengan metode *Latent Dirichlet Allocation* (LDA) untuk mengetahui tema dalam komentar By.U. LDA adalah model probabilitas untuk menganalisis topik dalam dokumen sebagai kombinasi dari topik tersembunyi [26]. Metode LDA digunakan untuk memproses data komentar dilakukan secara optimal dalam mengekstraksi setiap topik, memodelkan topik, menghasilkan istilah indeks dalam setiap cluster topik, dan visualisasi komputer dari topik [27].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Penelitian ini menggunakan data dari Instagram yang berisi teks bahasa Indonesia dan memproses dengan Python untuk *crawling* data. Data opini pengguna By.U diperoleh sebanyak 7408 dalam format *xlsx*. Tabel 2 merupakan contoh kumpulan data yang akan digunakan dalam penelitian ini.

Tabel 2. Dataset

no	Text
1	Jaringan nya tolong di stabil kan dong 😊
2	SIGNAL SAMPAAHHH....!!!!

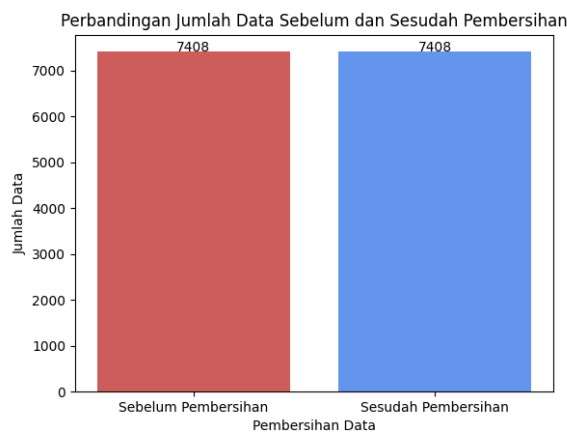
3.2 Text Pre-processing

Text Pre-processing adalah langkah selanjutnya setelah pengumpulan dataset untuk seleksi data teks lebih terstruktur dengan melakukan *cleaning* dan *case folding*. Proses *cleaning* dilakukan untuk menghilangkan karakter atau simbol seperti link URL, emoji, simbol, angka, baris kosong, kolom kosong, *missing value*, dan data duplikat. Pada Tabel 3. menunjukkan hasil data setelah proses *cleaning*.

Tabel 3. *Cleaning*

Sebelum melakukan <i>cleaning</i>	Sesudah melakukan <i>cleaning</i>
Jaringan nya tolong di stabil kan dong 😊	Jaringan nya tolong di stabil kan dong
SIGNAL SAMPAAHHH....!!!!	SIGNAL SAMPAAHHH

Pada Gambar 3. menunjukkan jumlah data sebelum dan sesudah pembersihan data. Kategori “Sebelum Pembersihan” dan “Sesudah Pembersihan” terletak di sumbu x, dan “Jumlah Data” terletak di sumbu y. Jumlah data sebelum dan sesudah pembersihan adalah sama, yaitu 7408 data, yang ditunjukkan dengan tinggi batang yang sama untuk kedua kategori dan angka di atas batang yang menunjukkan jumlah data.



Gambar 3. Perbandingan jumlah data

Tahap *case folding* dilakukan untuk mengubah sebuah teks dokumen ke dalam huruf kecil (*lowercase*). Dalam Tabel 4. menampilkan hasil setelah melakukan *case folding*.

Tabel 4. *Case folding*

Sebelum melakukan <i>case folding</i>	Sesudah melakukan <i>case folding</i>
Jaringan nya tolong di stabil kan dong 😊	jaringan nya tolong di stabil kan dong
SIGNAL SAMPAAHHH....!!!!	signal sampaaahhh

3.3 Text Processing

Text processing merupakan proses pengolahan teks dokumen menggunakan komputasi dengan tujuan mendapatkan informasi yang relevan. Adapun tahapan yang dilakukan *text processing*, yaitu *stemming*, *case folding*, *stopword removal*, frekuensi kata, keterkaitan kata, dan TF-IDF, sebagai berikut:

1) *Stemming*

Tahap *stemming* adalah sebuah teknik mengubah kata yang memiliki imbuhan ke dalam bentuk kata dasar. Pada Tabel 5. menunjukkan hasil setelah dilakukan *stemming*.

Tabel 5. *Stemming*

Sebelum melakukan <i>stemming</i>	Sesudah melakukan <i>stemming</i>
jaringan nya tolong di stabil kan dong	jaringan nya tolong di stabil kan dong
signal sampaaahhh	signal sampah

2) *Tokenization*

Setelah melakukan tahap *stemming*, kemudian tahap *tokenization* dilakukan untuk mengubah kalimat menjadi kata-kata. Pada Tabel 6. menunjukkan hasil setelah dilakukan *tokenization*.

Tabel 6. *Tokenization*

Sebelum melakukan <i>tokenization</i>	Sesudah melakukan <i>tokenization</i>
jaringan nya tolong di stabil kan dong	['jaring', 'nya', 'tolong', 'di', 'stabil', 'kan', 'dong']
signal sampah	['signal', 'sampah']

3) *Stopword Removal*

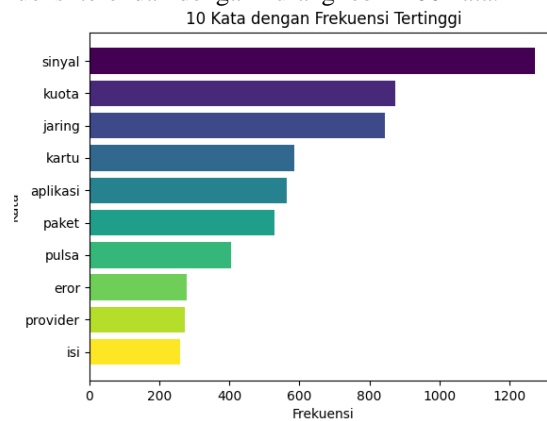
Selanjutnya langkah *stopword removal*, pada tahap ini melakukan penghapusan kata yang tidak umum dan tidak memiliki makna. Pada Tabel 7. menampilkan hasil setelah melakukan *stopword removal*.

Tabel 7. *Stopword removal*

Sebelum melakukan <i>stopword removal</i>	Sesudah melakukan <i>stopword removal</i>
jaring nya tolong di stabil kan dong	['jaring', 'stabil']
signal sampah	['signal']

4) Frekuensi Kata

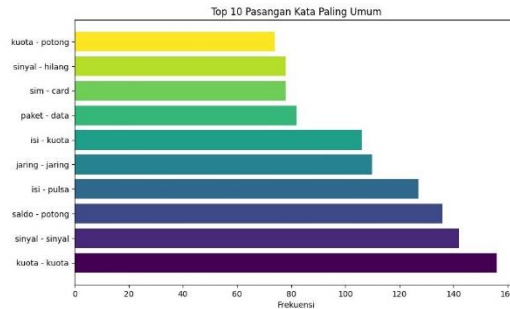
Selanjutnya tahap ini, dilakukan perhitungan jumlah kata yang sering muncul dalam teks dokumen dan melakukan identifikasi terhadap kata yang dapat menjadi kata kunci penting. Pada Gambar 4. merupakan hasil visualisasi menunjukkan barplot dengan warna berbeda untuk setiap kata dan frekuensi. Diagram ini terdiri dari dua sumbu: sumbu x menunjukkan “Frekuensi”, dan sumbu y menunjukkan kata-kata yang sering muncul dengan setiap barplot. Dalam diagram berikut menunjukkan bahwa kata sinyal memiliki frekuensi tertinggi dengan lebih dari 1000 kata, dan kata apk memiliki frekuensi terendah dengan kurang lebih 200 kata.



Gambar 4. Barplot frekuensi kata

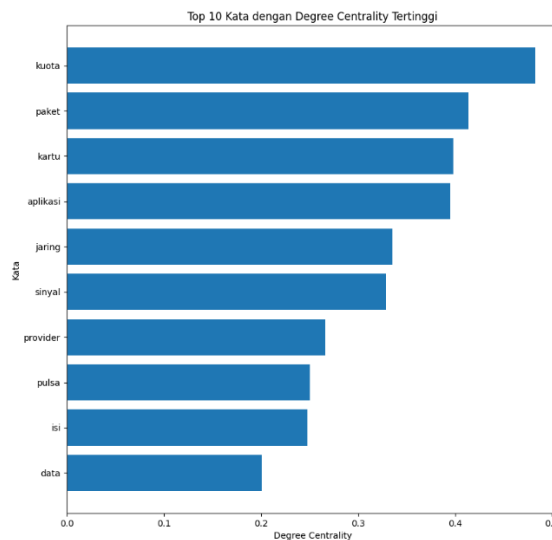
5) Keterkaitan Kata

Gambar 6 menunjukkan 10 pasangan kata yang paling umum digunakan dalam bahasa Indonesia, dan frekuensinya ditampilkan dalam plot horizontal berwarna. Sumbu y menunjukkan pasangan kata, dan sumbu x menunjukkan frekuensi mulai dari 0 hingga 160. Grafik ini memberikan ikhtisar kata-kata yang umum digunakan dalam kumpulan data atau konteks tertentu Pola yang benar.



Gambar 6. Barplot keterkaitan kata

Pada Gambar 7. diagram batang horizontal menunjukkan 10 kata dengan nilai "Degree Centrality" tertinggi. Dalam suatu jaringan, tingkat sentralitas adalah ukuran yang menunjukkan seberapa terhubungnya suatu node (dalam hal ini "kata") dengan node lainnya. Nilai sentralitas menunjukkan bahwa terdapat lebih banyak koneksi atau hubungan antar kata dalam jaringan.



Gambar 7. Degree centrality

Pada Gambar 7. sumbu y berisis daftar 10 kata, antara lain: kuota, paket, kartu, aplikasi, jaring, sinyal, provider, pulsa, isi, data. Kata-kata seperti "jaringan", "sinyal", "penyedia", "pulsa", "konten", dan "data" memiliki sentralitas yang lebih rendah namun tetap penting. Sentralitas kata "paket" relatif tinggi sekitar 0.4, sedangkan sentralitas "aplikasi" dan "kartu" sedikit lebih rendah dibandingkan "paket". Secara keseluruhan, diagram ini menunjukkan kata-kata dalam jaringan yang dianalisis yang paling terkait atau berdampak, dengan "kuota" menjadi kata yang paling penting.

6) TF-IDF (TermFrequency - Inverse Document Frequency)

Pada tahap ini, berbagai perhitungan dilakukan untuk menganalisis informasi kata, termasuk frekuensi kemunculan dan skor TF-IDF untuk setiap kata. Selain itu, frekuensi dokumen dapat dihitung berdasarkan jumlah dokumen yang mengandung kata tertentu. Kemudian, hasil TF dikalikan dengan hasil IDF. Hasil perhitungan TF-IDF disajikan pada Tabel 8.

Tabel 8. TF-IDF

no	Kata	Frekuensi	TF	IDF	TF-IDF
----	------	-----------	----	-----	--------

1	aktivasi	2.0	0.000013	6.640132	0.000084
2	admin	3.0	0.000019	6.640132	0.000125
3	akses	6.0	0.000038	6.640132	0.000251

3.4 Analisis Sentimen

Tahap menganalisis sentimen untuk melakukan klasifikasi teks ke dalam tiga kategori yaitu positif, negatif, atau netral. Adapun langkah yang akan dilakukan yaitu melakukan skoring dan labeling.

1) *Skoring*

Proses *skoring* dilakukan dengan mengukur teks atau menilai sentimen yang ada di dalamnya, dan jumlah skor dapat berkisar antara -6 hingga +7. Dalam melakukan *skoring* menggunakan dokumen kamus bahasa Indonesia yang sudah tersedia, kemudian dilakukan komputasi untuk mengukur nilai dalam teks. Hasil *skoring* dapat dilihat pada Tabel 9.

Tabel 9. Hasil *skoring*

no	Text	Score
1	jaring nya tolong di stabil kan dong	1
2	signal sampah	-1

2) *Labeling*

Setelah melakukan *skoring*, kemudian melakukan *labeling* berdasarkan skor tersebut. Skor yang berkisar antara -6 hingga -1 menandakan sentimen negatif, sedangkan skor antara 0 hingga 1 menunjukkan sentimen netral. Sebaliknya, skor 2 hingga 7 menandakan sentimen positif. Dalam pelabelan menggunakan dokumen kamus bahasa Indonesia yang tersedia, kemudian pemberian label dengan menggunakan komputasi. Hasil *labeling* dapat dilihat pada Tabel 10.

Tabel 10. Hasil *labeling*

no	Text	Score	Sentimen
1	jaring nya tolong di stabil kan dong	1	Netral
2	signal sampah	-1	Negatif

3.5 *K-Nearest Neighbors*

Pada tahap ini mengklasifikasikan sentimen menggunakan algoritma *K-Nearest Neighbors* untuk menentukan seberapa akurat model dalam mengklasifikasikan teks. Tahap-tahap Algoritma *K-Nearest Neighbors* antara lain, sebagai berikut:

1) Persiapan Data

Dataset yang diperoleh dibagi menjadi 80 persen data pelatihan dan 20 persen data pengujian menggunakan fungsi `train_test_split`. Data pelatihan terdiri dari atribut dan kelas yang telah melalui tahap *cleaning* dan diproses dengan benar. Kemudian, dilakukan pengujian pada performa kinerja model.

2) Menentukan Nilai K

Pada langkah ini nilai K diuji dari 1 hingga 25, untuk mencari K dengan error minimum. K=1 dengan minimum *error* 0.23 memberikan hasil terbaik, namun model dapat menjadi sensitif terhadap *noise* data (*overfitting*) saat nilai K terlalu kecil. Oleh karena itu, untuk mendapatkan nilai K yang optimal untuk generalisasi yang lebih baik, sebaiknya melakukan *cross-validation*.

Setelah melakukan pengujian dengan menggunakan teknik *k-fold cross-validation*, untuk memastikan nilai K yang optimal. Nilai optimal K=23 untuk dataset yang diuji, dengan akurasi rata-rata tertinggi yaitu 0.37. Oleh karena itu, untuk memastikan bahwa model yang paling akurat dan stabil, disarankan untuk mempertimbangkan tren keseluruhan dan menguji beberapa nilai K di sekitar nilai optimal. Hasil pengujian menunjukkan bahwa nilai K optimal yaitu K=23 dengan akurasi terbaik sebesar 0.37.

3) Menghitung Jarak

Pada tahap ini dapat menggunakan *Euclidean distance* untuk menghitung jarak antara titik data dalam dataset pelatihan dan titik data dalam dataset pengujian. Tabel 11. menunjukkan keduanya memiliki nilai yang sama, jarak antara "Data Uji 1" dan "Data Pelatihan 1" adalah nol.

Tabel 11. Hasil perhitungan jarak

Data Uji/Data Pelatihan	Data Pelatihan 1	Data Pelatihan 2	Data Pelatihan 3	Data Pelatihan 4	Data Pelatihan 5
Data Uji 1	2147.00069865	2570.00175097	2594.00173477	1973.00228079	2986.00050234
Data Uji 2	89.03370148	334.02095743	4830.00144928	263.00760445	750.00399999
Data Uji 3	4286.00128325	4709.00222977	455.02307634	4112.0001216	5125.00107317
Data Uji 4	741.00000000	1164.00085911	4000.00025	567.01587279	1580.00000000
Data Uji 5	853.00351699	1276.00548588	3888.00180041	679.0029455	1692.00177305

Penggunaan rumus *Euclidean* untuk melakukan perhitungan jarak antara "Data Uji" dan "Data Pelatihan" untuk menentukan tetangga terdekat dalam metode *K-Nearest Neighbors* (KNN).

4) Menentukan Tetangga Terdekat

Pada langkah ini, menghitung tingkat kemiripan antara dokumen dengan indeks tetangga terdekat menggunakan nilai $K = 23$ untuk mencari kalimat serupa. Tabel 12. menunjukkan tetangga terdekat menurut komentar pengguna By.U.

Tabel 12. Hasil menentukan tetangga terdekat

No	Kalimat	Sentimen	Tetangga Terdekat
1	jaring lot edge tiap sat	Netral	['sinyal lot', 'masih lot byu rasa edge']
2	koneksi internet nya makin kesini malah makin lot	Negatif	['ampun makin hari makin ga jelas sinyal', 'makin lama makin ancur ini byu']
3	tetap nyala abang bareng yang bikin aman jaya unlimited mbps	Positif	['byu makin kesini aman jaya banget kuota mbps rb spednya gila banget bintang sih', 'penguna setia byuid pergi mana pun sinyal tetap aman']

5) Menentukan Kelas

Pada Tabel 13. menunjukkan hasil perbandingan sentimen asli dan prediksi model analisis sentimen pada tiga contoh kalimat dokumen untuk mengevaluasi kinerja model.

Tabel 13. Hasil menentukan kelas

Kalimat	Sentimen	Prediksi Sentimen
provider panteq buat vc aja sinyal ampun	Netral	Negatif
ini kenapa dah kmaren gw beli kuota kaget hari masa bilang besok hangus blom hari loh	Negatif	Netral
sinyal kok jumping banget ya ganyaman banget makenya begini	Positif	Netral

3.6 Evaluasi

Berdasarkan hasil pengujian metode *K-Nearest Neighbors* menunjukkan akurasi 0.73, presisi 0.76 dalam mengklasifikasikan sentimen negatif, *recall* 0.39 untuk menemukan semua kasus negatif, dan *f1-score* 0.52. Pada sentimen netral, presisi 0.72, *recall* 0.94, dan *f1-score* 0.81. Namun, model tidak dapat mengklasifikasikan sentimen positif dengan baik. Hasil evaluasi dapat dilihat pada Tabel 14.

Tabel 14. Hasil evaluasi

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
Negatif	0.76	0.39	0.52	494
Netral	0.72	0.94	0.81	941
Positif	0.00	0.00	0.00	46
accuracy			0.73	1481
macro avg	0.49	0.44	0.44	1481
weighted avg	0.71	0.73	0.69	1481

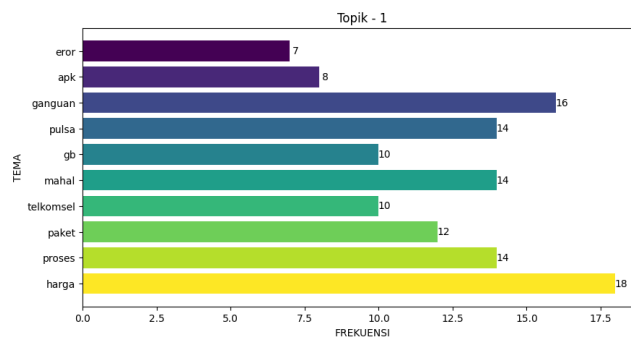
Tabel 15 menunjukkan hasil evaluasi model klasifikasi dengan tiga kelas: *Class 0*, *Class 1*, dan *Class 2*. Hasil evaluasi mencakup jumlah *False Negatif* (FN), *True Positif* (TP), *True Negatif* (TN), dan *False Positif* (FP) untuk setiap kelas.

Tabel 15. *Confusion matrix*

<i>Class</i>	TP	FP	FN	TN
0	193	60	301	927
1	882	346	59	194
2	0	0	46	1435

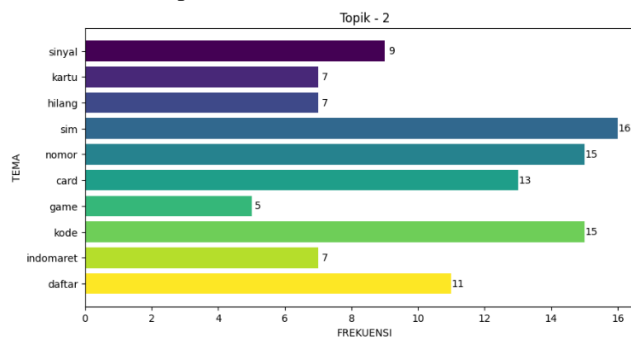
Setelah menguji dataset, dilakukan *topic modelling* menggunakan metode *Latent Dirichlet Allocation* (LDA) untuk mengetahui topik apa saja yang dibahas pada komentar pengguna By.U. Pada Gambar 11 hingga Gambar 15 menunjukkan bahwa diagram batang horizontal menampilkan frekuensi berbagai tema dengan topik yang berbeda. Diagram ini memiliki delapan batang berwarna berbeda yang masing-masing mewakili kategori tertentu pada sumbu y, dan sumbu x menunjukkan frekuensi tersebut.

Gambar 11. menampilkan bahwa diagram batang horizontal dengan judul “Topik – 1”. Kategori dari atas ke bawah adalah error, gangguan, apk, pulsa, gb, mahal, telkomsel, paket, proses, harga, dan telkomsel. Gambar ini menarik karena menggambarkan data telekomunikasi seperti tema "harga" memiliki frekuensi tertinggi.



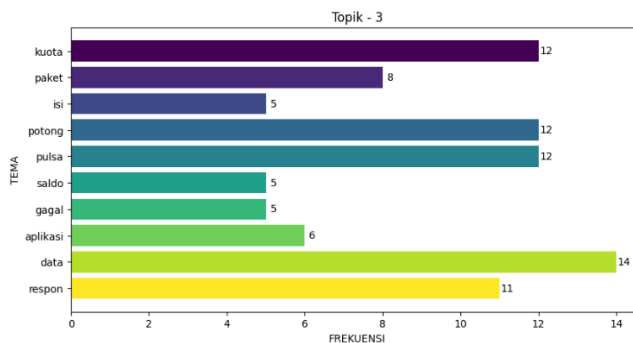
Gambar 11. Topik 1

Gambar 12. menampilkan bahwa diagram batang horizontal dengan judul “Topik – 2”. Kategori dari atas ke bawah adalah sinyal, hilang, kartu, SIM, nomor, kode, game, dan Indomaret. Gambar ini menarik karena menggambarkan pendaftaran *SIM card* seperti tema "SIM" memiliki frekuensi tertinggi.



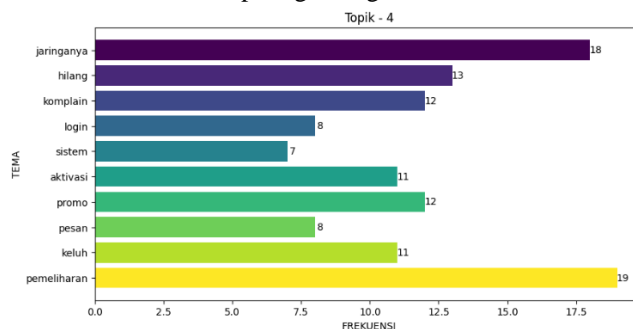
Gambar 12. Topik 2

Gambar 13. menampilkan bahwa diagram batang horizontal menampilkan topik “Topik – 3” yang membahas masalah paket data, aplikasi, transaksi, dengan tema "data" memiliki frekuensi tertinggi.



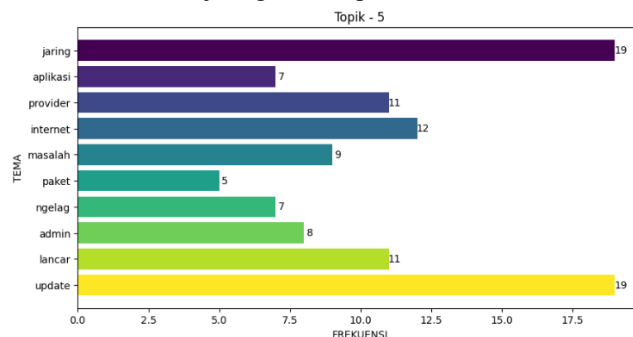
Gambar 13. Topik 3

Gambar 14. menampilkan bahwa diagram batang horizontal berjudul “Topik – 4” yang menunjukkan berbagai masalah teknis. Tema “pemeliharaan” muncul paling sering dan memiliki frekuensi tertinggi.



Gambar 14. Topik 4

Gambar 15. menampilkan bahwa diagram batang horizontal berjudul “Topik – 5” yang membahas mengenai respon admin dan jaringan internet. Tema “jaring” dan “update” memiliki frekuensi yang sama tinggi.



Gambar 15. Topik 5

Diagram pada Gambar 11 hingga 15 secara visual dapat mewakili data frekuensi untuk berbagai tema. Hasil *topic modelling* dapat berguna untuk menganalisis tren atau pola dalam konteks tertentu, seperti pertanyaan layanan pelanggan atau tema umpan balik.

4. KESIMPULAN

Berdasarkan hasil penelitian bahwa nilai akurasi menggunakan metode *K-Nearest Neighbors* (KKN) adalah 73 persen pada 7408 dataset. Tema yang dibahas dalam komentar pengguna By.U meliputi data telekomunikasi, pendaftaran sim *card*, masalah paket internet, aplikasi, transaksi, keluhan teknis, respon admin, dan jaringan internet. Model kesulitan memprediksi sentimen dengan akurasi yang rendah untuk kelas positif, netral, atau negatif. Diperlukan peningkatan model dengan sampel yang lebih besar atau fitur yang lebih baik. Hasil *topic modelling* dapat membantu menganalisis tren layanan pelanggan atau tema umpan balik. Hasil penelitian ini dapat memberikan wawasan bagi penyedia layanan untuk meningkatkan kualitas layanan. Sehingga pengguna By.U dapat menikmati kualitas layanan dengan harga terjangkau.

REFERENSI

- [1] J. Iskandar, V. C. Mawardi, and ..., "Analisis Sosial Media Layanan Provider Internet Menggunakan Algoritma XGBOOST," ... *Nas. CORIS 2022*, 2022, [Online]. Available: <https://corisindo.stikom-bali.ac.id/penelitian/index.php/semnas/article/view/11>
- [2] I. Verawati and B. S. Audit, "Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1411, 2022, doi: 10.30865/mib.v6i3.4132.
- [3] F. A. Anggari, "Analisis Kesiapan Dan Penerimaan Aplikasi All-Digital Provider BY.U Dengan Menggunakan Technology Readiness And Acceptance Model (TRAM)," *Skripsi*, no. 8.5.2017, pp. 2003–2005, 2022.
- [4] Telkomsel, "Telkomsel Luncurkan by.U, Layanan Selular Prabayar Digital End-to-end Pertama di Indonesia," <https://www.telkomsel.com/about-us/news/telkomsel-luncurkan-byu-layanan-selular-prabayar-digital-end-end-pertama-di-indonesia>, 2019.
- [5] T. Ajeng, "6 Hal yang Bikin Kamu Wajib 'Pindah' ke by.U! Artikel ini telah tayang di Idntimes.com dengan judul '6 Hal yang Bikin Kamu Wajib "Pindah" ke by.U!,'" *IDN Times*, 2024. <https://www.idntimes.com/tech/trend/tisa-ajeng/pindah-ke-byu-csc?page=all>
- [6] L. Utama, "by.U Raih 20 Juta Unduhan Aplikasi, Mayoritas Gen Z," *viva.ci.id*, 2023. <https://www.viva.co.id/amp/digital/digilife/1647082-by-u-raih-20-juta-unduh-an-aplikasi-mayoritas-gen-z>
- [7] R. M. Candra and A. N. Rozana, "Klasifikasi Komentar Bullying pada Instagram Menggunakan Metode K-Nearest Neighbor," vol. 5, no. 1, pp. 45–52, 2020.
- [8] A. S. Neogi, K. A. Garg, R. K. Mishra, and Y. K. Dwivedi, "Sentiment analysis and classification of Indian farmers' protest using twitter data," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, p. 100019, 2021, doi: 10.1016/j.jjime.2021.100019.
- [9] B. H. Agtira, H. H. Handayani, and A. F. N. Masruriyah, "Perbandingan Algoritma NBC dan Decision Tree pada Sentimen Analisis Mengenai Vaksinasi Covid-19 Di Indonesia," *Remik*, vol. 7, no. 1, pp. 704–712, 2023, doi: 10.33395/remik.v7i1.12151.
- [10] L. Mamluatul Hikmah, N. Rahaningsih, and R. Danar Dana, "Penerapan Algoritma K-Nearest Neighbor Pada Twitter Untuk Analisis Sentimen Masyarakat Terhadap Lgbt Di Indonesia," *J. Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 284–291, 2024, [Online]. Available: <https://databoks.katadata.co.id/>,
- [11] H. Prasetyo and Nurapriani, "Jurnal KomtekInfo Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan," vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [12] A. Firyal, L. Ramadhina, and E. Sofian, "Perbandingan Metode K-Nearest Neighbor dan Pohon Keputusan dalam Analisis Sentimen Data Ulasan Aplikasi Pinjaman Online Berizin OJK Di Google Play," vol. VII, pp. 115–124, 2024.
- [13] H. Bryan, R. Intan, and H. Juwiantho, "Pemanfaatan Text Summarization Dengan Support Vector Machine Dan K-Nearest Neighbor Pada Analisis Sentimen Untuk Mempermudah Pengguna Membaca Review Game Steam," *J. Infra*, vol. 10, no. 1, pp. 31–36, 2022.
- [14] V. F. Rusdhi and I. Sari, "Identifikasi Topik Artikel Berita Menggunakan Topic Modelling Dengan Latent Dirichlet Allocation," *J. Ilm. Inform. Komput.*, vol. 27, no. 2, pp. 169–176, 2022, doi: 10.35760/ik.2022.v27i2.6829.
- [15] R. C. Putri, "Analisis sentimen provider by. u dengan metode naïve bayes classifier dan k-nearest neighbor (KNN)," *Repository.Uinjkt.Ac.Id*, 2023, [Online]. Available: <https://repository.uinjkt.ac.id/dspace/handle/123456789/76502>
- [16] M. F. Asshiddiqi and K. M. Lhaksana, "Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI," vol. 7, no. D, pp. 2019–2020, 2020.
- [17] C. M. Annur, "Jumlah Pengguna Instagram Indonesia Terbanyak ke-4 di Dunia," *Databoks.Katadata*, 2023. <https://databoks.katadata.co.id/datapublish/2023/05/04/jumlah-pengguna-instagram-indonesia-terbanyak-ke-4-di-dunia>
- [18] I. T. Julianto, "Analisis Sentimen Terhadap Sistem Informasi Akademik Institut Teknologi Garut," *J. Algoritm.*, vol. 19, no. 1, pp. 449–456, 2022, doi: 10.33364/algoritma/v.19-1.1112.

- [19] A. Siregar, S. Faisal, T. Tukino, A. Puspabhuana, and M. Simarangkir, “Comparison Study Of Term Weighting Optimally With SVM In Sentiment Analysis,” 2019, doi: 10.4108/eai.18-7-2019.2288508.
- [20] D. Remaldi, D. Wahiddin, and Y. Cahyana, “Identifikasi Kesegaran Ikan Nila Berdasarkan Warna Insang Menggunakan Algoritma K-Nearest Neighbor (K-NN),” *Sci. Student J. Information, Technol. Sci.*, vol. II, no. 1, pp. 197–202, 2021, [Online]. Available: <http://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/download/247/175>
- [21] F. Aryanto, A. Fauzi, A. Fitri Nur Masruriyah, A. Lia Hananto, and Darmansyah, “Sentiment Analysis Of Vaccination Using The K-Nearest Neighbor Algorithm,” *Edutran Comput. Sci. Inf. Technol.*, vol. 1, no. 1, pp. 34–41, 2023, doi: 10.59805/ecsit.v1i1.6.
- [22] Q. A. A’yuniyah and M. Reza, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru,” *Indones. J. Inform. Res. Softw. Eng.*, vol. 3, no. 1, pp. 39–45, 2023, doi: 10.57152/ijirse.v3i1.484.
- [23] T. A.M and A. Yaqin, “Perbandingan Algoritma Naïve Bayes, K-Nearest Neighbors dan Random Forest untuk Klasifikasi Sentimen Terhadap BPJS Kesehatan pada Media Twitter,” *InComTech J. Telekomun. dan Komput.*, vol. 12, no. 1, p. 01, 2022, doi: 10.22441/incomtech.v12i1.13642.
- [24] F. D. Ananda and Y. Pristyanto, “Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 407–416, 2021, doi: 10.30812/matrik.v20i2.1130.
- [25] E. Hasibuan and E. A. Heriyanto, “Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier,” *J. Tek. dan Sci.*, vol. 1, no. 3, pp. 13–24, 2022, doi: 10.56127/jts.v1i3.434.
- [26] Y. Kustyaningsih, Y. Permana, P. Studi, S. Informasi, U. Trunojoyo, and J. Timur, “Penggunaan Latent Dirichlet Allocation (LDA) dan Support- Vector Machine (SVM) Untuk Menganalisis Sentimen Berdasarkan Aspek Dalam Ulasan Aplikasi EdLink The Use of Latent Dirichlet Allocation (LDA) and Support-Vector Machine (SVM) to Analyze Sent,” vol. 13, no. 1, pp. 127–136, 2024, doi: 10.34148/teknika.v13i1.746.
- [27] A. F. N. Masruriyah, C. E. Sukmawati, and H. Y. Novita, “Pengelompokan Topik Cuitan Pengguna Twitter Terhadap Kuliah Kerja Nyata (Kkn) Di Indonesia Menggunakan Latent Dirichlet Allocation,” pp. 1128–1133, 2022.