

Analisis *Pre-processing* Sentimen Terhadap Komentar Layanan Indihome pada Twitter

Sentiment Pre-processing Analysis of Indihome Service Comments on Twitter

Achmad Novanto^{a,1}, Dolly Indra^{b,2}, dan Wistiani Astuti^{a,3}

^aTeknik Informatika, Universitas Muslim Indonesia, Makassar, Indonesia

^bSistem Informasi, Universitas Muslim Indonesia, Makassar, Indonesia

¹13020180039@umi.ac.id; ²dolly.indra@umi.ac.id; ³wistiani.astuti@umi.ac.id

*corresponding author

Informasi Artikel	ABSTRAK
<p>Diserahkan : 2 Februari 2024 Diterima : 28 Februari 2024 Direvisi : 29 Februari 2024 Diterbitkan : 29 Februari 2024</p> <p>Kata Kunci: Analisis Sentimen Indihome Twitter Preprocessing</p>	<p>Perusahaan telekomunikasi seperti PT Telkom Indonesia dengan layanannya, IndiHome, memanfaatkan kemajuan teknologi untuk menyediakan layanan digital berbasis Internet, Telepon Rumah, dan TV Interaktif/IPTV. Meskipun sudah menjangkau seluruh Indonesia, pemahaman mengenai kepuasan pengguna terhadap layanan IndiHome masih perlu diperdalam. Penelitian ini difokuskan pada analisis sentimen pengguna terhadap layanan IndiHome melalui media sosial twitter. Twitter menjadi platform yang signifikan dalam mengekspresikan pandangan, kritik, dan kepuasan pengguna. Pembatasan karakter dalam setiap cuitan memunculkan gaya bahasa baru, yang memicu kreativitas pengguna. Meski demikian, menganalisis sentimen dari <i>tweet</i> memiliki tantangan tersendiri, terutama karena penggunaan kata-kata non-baku dan bahasa informal. Oleh karena itu, pentingnya <i>preprocessing</i> data dalam analisis sentimen menjadi fokus utama penelitian ini. Langkah awal dalam penelitian ini bertujuan untuk meningkatkan keberhasilan klasifikasi sentimen dengan membersihkan dan normalisasi data <i>tweet</i>. Hasil penelitian diharapkan dapat memberikan wawasan yang lebih akurat mengenai respons pengguna terhadap layanan IndiHome. Melalui langkah-langkah <i>preprocessing</i> yang dilibatkan, penelitian ini menyimpulkan bahwa data yang telah dipersiapkan menjadi lebih siap untuk tahap analisis sentimen. Dengan demikian, analisis sentimen dapat memberikan hasil yang lebih relevan dan akurat, membuka peluang untuk mengambil langkah-langkah strategis terkait dengan polarisasi sentimen yang teridentifikasi.</p>
<p>Keywords: Sentiment Analysis Indihome Twitter Preprocessing</p>	<p>ABSTRACT</p> <p><i>Telecommunications companies such as PT Telkom Indonesia with its service, IndiHome, utilize technological advances to provide digital services based on the Internet, Home Telephone and Interactive TV/IPTV. Even though it has reached all of Indonesia, understanding of user satisfaction with IndiHome services still needs to be deepened. This research focuses on analyzing user sentiment towards IndiHome services via Twitter social media. Twitter has become a significant platform for expressing views, criticism and user satisfaction. The character limit in each tweet gives rise to new language styles, which triggers user creativity. However, analyzing sentiment from tweets has its own challenges, especially due to the use of non-standard words and informal language. Therefore, the importance of data preprocessing in sentiment analysis is the main focus of this research. The initial step in this research aims to increase the success of sentiment classification by cleaning and normalizing tweet data. It is hoped that the research results will provide more accurate insight into user responses to IndiHome services. Through the preprocessing steps involved, this research concludes that the data that has been prepared is better prepared for the sentiment analysis stage. Thus, sentiment analysis can provide more relevant and accurate results, opening up opportunities to take strategic steps related to the identified sentiment polarization.</i></p>

This is an open access article under the

[CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



I. Pendahuluan

Dalam era globalisasi saat ini, hampir semua aktivitas manusia memberikan perhatian kepada perkembangan teknologi informasi. Berkat kemajuan dalam teknologi komunikasi dan informasi, hubungan antara negara, organisasi, dan individu. Selain itu, orang-orang lebih banyak mengetahui tentang ilmu pengetahuan, di mana informasi memiliki peran penting dan nyata dalam kehidupan di era saat ini. Oleh karena itu, masyarakat harus lebih sadar akan teknologi untuk mengikuti perkembangan saat ini [1].

Hal ini menjadi peluang besar bagi Perusahaan telekomunikasi dalam jasa Internet yaitu PT Telkom Indonesia dengan produknya yaitu IndiHome. IndiHome merupakan layanan digital yang menyediakan Internet, Telepon Rumah dan TV Interaktif/IPTV (IndiHome TV) dengan beragam pilihan paket serta layanan tambahan yang bisa dipilih sesuai kebutuhan Anda. Saat ini, jaringan IndiHome sudah tersebar di seluruh wilayah Indonesia, dan terus berinovasi untuk memenuhi kebutuhan internet yang lebih baik bagi Masyarakat[2].

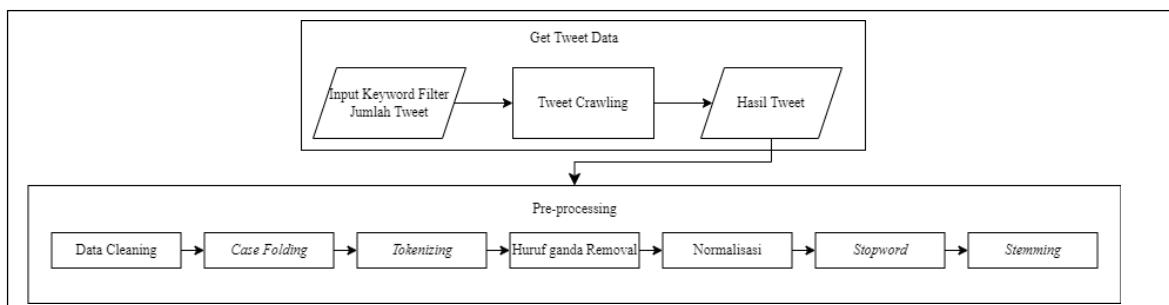
Indihome menyediakan layanan melalui offline maupun online untuk kepuasan pelanggan. Namun, pengguna merasa bingung ketika melihat bagaimana performa layanan, mulai dari kepuasan, kritik, hingga ketidakpuasan, sehingga data dari pengguna tidak terlihat. Oleh karena itu, PT Telkom Indonesia membuka situs media sosial adalah twitter. Media sosial twitter memiliki peranan penting dalam kegiatan penyaluran informasi dan komunikasi. Indonesia termasuk salah satu bagian dari pertumbuhan pengguna aktif terbesar menurut laporan finansial twitter kuartal ke-3 tahun 2019, suatu platform untuk berbicara dan mengaktualisasikan diri[3], [4].

Twitter adalah platform media sosial yang sangat populer di internet. Dengan menggunakan twitter, pengguna dapat berbagi pendapat, mengikuti acara, dan berinteraksi dengan pengguna lainnya. Pertumbuhan penggunaan media sosial membuka peluang baru untuk menganalisis berbagai aspek dan pola dalam komunikasi. Contohnya, data media sosial dapat dianalisis untuk mendapatkan wawasan tentang isu-isu terkini, tren, tokoh berpengaruh, dan jenis informasi lainnya[5]. twitter menerapkan pembatasan jumlah karakter sebanyak 280 dalam setiap cuitan atau unggahan, baik itu huruf atau karakter. Oleh karena itu, pengguna diharuskan untuk merinci dan merangkum pengungkapannya. Dampak dari pembatasan ini adalah kemungkinan munculnya kosakata baru, karena kreativitas masyarakat dalam menggunakan bahasa Indonesia di media sosial meningkat[6].

Saat menganalisis sentimen dengan menggunakan data twitter, ada beberapa permasalahan yang mungkin timbul. Pada umumnya, tweet masih mengandung banyak kata-kata non-baku, seperti penulisan singkat dan penggunaan bahasa informal[7]. Oleh karena itu, perlu dilakukan preprocessing terhadap data tweet sebagai tahapan awal dalam analisis sentimen sehingga tepat menghasilkan bentuk data dan meningkatkan keberhasilan klasifikasi[8]. Selain itu, Analisis sentimen digunakan untuk mengamati respons pengguna pada pelayanan tertentu. Metodenya melibatkan klasifikasi untuk menentukan apakah suatu kalimat, dokumen, atau pendapat mengandung polarisasi tertentu dalam teks. Hal ini bertujuan agar langkah-langkah strategis dapat diambil di masa mendatang terkait dengan polarisasi yang teridentifikasi[9]. Penelitian ini bertujuan untuk mencari tahu pengaruh dari penerapan teknik *preprocessing* yang digunakan.

II. Metode

Pada penelitian ini, ada tahapan dalam melakukan analisis sentimen terhadap komentar pada Pelayanan Indihome di media sosial twitter seperti Gambar 1.



Gambar 1 Tahapan Pre-processing

A. Pengumpulan tweet data

Pada tahapan ini menunjukkan proses pengambilan data Dimana proses untuk mendapat sampel pada data. Proses ini dilakukan mengambil data *tweet* dengan menginput *Keyword Filter* untuk menentukan kata kunci yang menyaring pada *tweet*. Keyword yang akan diambil adalah *indihome* atau *indihomecare*. Selanjutnya, *tweet crawling* yang digunakan adalah *tweet harvest* dalam bahasa pemrograman *python*[10]. Setelah *crawling* maka akan menjadi dataset dan data disimpan dalam bentuk file *Comma Separated Values (CSV)*. Hasil *crawling* data dapat terlihat pada Gambar 2.

Gambar 2. Hasil *crawling* data

B. *Pre-processing*

Pada tahap *pre-processing* ada beberapa langkah yang dapat dilakukan *pre-processing* yaitu sebagai berikut:

1) *Data Cleaning*

Data Cleaning merupakan proses untuk membersihkan data yang bersifat *noise* seperti *symbol*, numerik, *emoji*, *mention* pada *tweet* dan *Uniform Resource Locator (URL)*.

2) *Case folding*

Case folding adalah suatu proses untuk mengubah semua huruf dalam suatu teks menjadi huruf kecil secara keseluruhan, atau disebut juga *lowercase*[11]. Contoh proses *case folding* dapat tercantum pada Tabel 1.

Tabel 1. Contoh *Case Folding*

Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
Padahal gua nggak puas banget dengan internet IndiHome	padahal gua nggak puas banget dengan internet indihome
Wahhh, bagus sekali pelayanannya makasih	wahhh bagus sekali pelayanannya makasih
kok, internetnya lemot bener. Tolong, perbaiki	kok internetnya lemot bener tolong, perbaiki

3) *Tokenizing*

Tokenizing adalah langkah di mana sekumpulan kata yang membentuk sebuah kalimat akan dipecah menjadi kata-kata tunggal atau token[12],[13]. Menggunakan *Tokenizing* mampu membedakan antara pemisah kata dan non-kata. Contoh proses *tokenizing* dapat tercantum pada Tabel 2.

Tabel 2. Contoh *Tokenizing*

Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
padahal gua nggak puas banget dengan internet indihome	['padahal', 'gua', 'nggak', 'puas', 'banget', 'dengan', 'internet', 'indihome']
wahhh, bagus sekali pelayanannya makasih	['wahhh', 'bagus', 'sekali', 'pelayanannya', 'makasih']
kok, internetnya lemot bener. tolong, perbaiki	['kok', 'internetnya', 'lemot', 'bener', 'tolong', 'perbaiki']

4) *Hapus huruf ganda*

Hapus huruf Ganda dimana tahapan untuk mengnormalkan huruf berulang. Contoh proses hapus huruf ganda dapat tercantum pada Tabel 3.

Tabel 3. Contoh Hapus Huruf Ganda

Sebelum Hapus Huruf Ganda	Setelah Hapus Huruf Ganda
['padahal', 'gua', 'nggak', 'puas', 'banget' 'dengan', 'internet', 'indihome']	['padahal', 'gua', 'ngak', 'puas', 'banget' 'dengan', 'internet', 'indihome']
['wahhh', 'bagus', 'sekali', 'pelayanannya', 'makasih']	['wah', 'bagus', 'sekali', 'pelayanannya', 'makasih']
['kok', 'internetnya', 'lemot', 'bener', 'tolong', 'perbaiki']	['kok', 'internetnya', 'lemot', 'bener', 'tolong', 'perbaiki']

5) *Normalisasi*

Normalisasi yaitu tahapan yang merubah kata tidak valid, kata tidak baku, atau kata singkatan menjadi kata baku[14]. Misalnya, pada kata 'gua' yang tidak baku, dapat diubah menjadi kata baku seperti 'saya'. Berikut contoh proses normalisasi pada Tabel 4.

Tabel 4. Contoh Normalisasi

Sebelum Normalisasi	Setelah Normalisasi
['padahal', 'gua', 'ngak', 'puas', 'banget' 'dengan', 'internet', 'indihome']	['padahal', 'saya', 'tidak', 'puas', 'sangat' 'dengan', 'internet', 'indihome']
['wah', 'bagus', 'sekali', 'pelayanannya', 'makasih']	['wah', 'bagus', 'sekali', 'pelayanannya', 'terima', 'kasih']
['kok', 'internetnya', 'lemot', 'bener', 'tolong', 'perbaiki']	['kenapa', 'internetnya', 'lambat', 'benar', 'tolong', 'perbaiki']

6) *Stopword*

Stopword adalah kata-kata yang sering muncul. Dalam *Information Retrieval* dan *text mining*, *stopwords* dianggap kurang berguna karena tidak memberikan informasi yang signifikan. Contohnya, kata-kata seperti "yang", "di", "pada", dan sejenisnya. Tujuan utama dari penghapusan *stopwords* adalah untuk mencegah mereka agar tidak mempengaruhi hasil pada tahap selanjutnya dari proses tersebut[15]. Berikut contoh proses *stopword* pada Tabel 5.

Tabel 5. Contoh *Stopword*

Sebelum <i>Stopword</i>	Setelah <i>Stopword</i>
['padahal', 'saya', 'tidak', 'puas', 'sangat' 'dengan', 'internet', 'indihome']	['padahal', 'saya', 'tidak', 'puas', 'sangat', 'internet', 'indihome']
['wah', 'bagus', 'sekali', 'pelayanannya', 'terima', 'kasih']	['wah', 'bagus', 'sekali', 'pelayanannya', 'terima', 'kasih']
['kenapa', 'internetnya', 'lambat', 'benar', 'tolong', 'perbaiki']	['kenapa', 'internetnya', 'lambat', 'benar', 'tolong', 'perbaiki']

7) *Stemming*

Stemming merupakan tahap di mana kata-kata dengan imbuhan diubah menjadi bentuk dasarnya[16]. Misal, dalam kata "memulai" diubah menjadi kata "mulai". Adapun contoh proses *stemming* dapat tercantum pada Tabel 6.

Tabel 6. Contoh *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
['padahal', 'saya', 'tidak', 'puas', 'sangat' 'dengan', 'internet', 'indihome']	['padahal', 'saya', 'tidak', 'puas', 'sangat', 'internet', 'indihome']
['wah', 'bagus', 'sekali', 'pelayanannya', 'terima', 'kasih']	['wah', 'bagus', 'sekali', 'layanannya', 'terima', 'kasih']
['kenapa', 'internetnya', 'lambat', 'benar', 'tolong', 'perbaiki']	['kenapa', 'internet', 'lambat', 'benar', 'tolong', 'perbaiki']

III. Hasil dan Pembahasan

Pada penelitian ini dimulai dengan *crawling* data twitter untuk mengumpulkan seluruh dataset dari komentar twitter. Hasil sampel pada dataset terangkum pada Tabel 7.

Tabel 7. Sampel dataset

Index	Data Tweet
1	@noblnolyf Hai, Kak Agar Dya bantu lebih lanjut. Infoin atas nama, nomor layanan dan alamat lewat DM ya. Kebijakan privasi layanan IndiHome bisa di cek melalui https://t.co/8xlv5hcUcB . Makasih -Dya
2	Indihome kenapa ya modemnya padahal indikator di lampunya normal2 aja tapi waktu mau connect gak bisa dan jaringannya nggak ada di hp @IndiHomeCare
3	@IndiHomeCare @IndiHome @YLKI_ID SEKARANG MAU TELEPON KELUAR MALAH RESPONNYA DIBLOKIR KARENA BELOM BAYAR, GUE UDAH BAYAR TAGIHAN DARI KAPAN TAU!!! BUKA ISOLIR TELEPONNYA SEKARANG!
4	@deddysat Kak Deddy, maaf ya bikin gak nyaman. Agar Dya bantu lebih lanjut. Infoin atas nama, nomor layanan dan alamat lewat DM ya. Kebijakan privasi layanan IndiHome bisa di cek melalui https://t.co/8xlv5hcUcB . Makasih -Dya
5	@YBudyyy Hai, Kak Agar Dya bantu lebih lanjut. Infoin atas nama, nomor layanan dan alamat lewat DM ya. Kebijakan privasi layanan IndiHome bisa di cek melalui https://t.co/8xlv5hcUcB . Makasih -Dya
6	@LeonaCallistaa Baik jika masih berkendala. Infoin atas nama, nomor layanan dan alamat lewat DM ya. Kebijakan privasi layanan IndiHome bisa di cek melalui https://t.co/8xlv5hcUcB . Makasih -Dya
7	@ArifirmanH @IndiHome Hai, Kak Arif Agar Dya bantu lebih lanjut. Infoin atas nama, nomor layanan dan alamat lewat DM ya. Kebijakan privasi layanan IndiHome bisa di cek melalui https://t.co/8xlv5hcUcB . Makasih -Dya
...	...
500	@IndiHomeCare siap kak nanti saya DM ðŸ™• btw kaka ga mau aku jadi brand ambassador indihome ? ðŸœŸðŸœŸ

Bagian ini akan melibatkan pemrosesan dataset dengan langkah pembersihan, seperti menghilangkan *mention/username*, *hashtag*, *emoji*, angka, symbol dan URL. Proses *text mining* akan diterapkan setelah selesai membersihkan kata-kata yang tidak diperlukan. Setelah proses pembersihan selesai, langkah selanjutnya adalah melakukan *case folding* untuk mengubah seluruh dataset menjadi huruf kecil. Pada tabel 8 menampilkan proses hasil dari *case folding*.

Tabel 8. Hasil *case folding*

Data
hai kak agar dya bantu lebih lanjut infoin atas nama nomor layanan dan alamat lewat dm ya kebijakan privasi layanan indihome bisa di cek melalui makasih dya
indihome kenapa ya modemnya padahal indikator di lampunya normal2 aja tapi waktu mau connect gak bisa dan jaringannya nggak ada di hp
sekarang mau telepon keluar malah responnya diblokir karena belum bayar gue udah bayar tagihan dari kapan tau buka isolir teleponnya sekarang
kak deddy maaf ya bikin gak nyaman agar dya bantu lebih lanjut infoin atas nama nomor layanan dan alamat lewat dm ya kebijakan privasi layanan indihome bisa di cek melalui makasih dya
hai kak agar dya bantu lebih lanjut infoin atas nama nomor layanan dan alamat lewat dm ya kebijakan privasi layanan indihome bisa di cek melalui makasih dya
...
siap kak nanti saya dm btw kaka ga mau aku jadi brand ambassador indihome

Langkah berikutnya melibatkan tokenisasi untuk mengambil kata-kata dari suatu kalimat dengan memisahkan kalimat menjadi kata-kata dan menghapus tanda baca. Setelah itu, dilakukan proses penghapusan Huruf Ganda untuk mengeliminasi huruf yang sama. Dari hasil *tokenizing* dan Hapus huruf Ganda terangkum pada Tabel 9.

Tabel 9. Hasil *tokenizing* dan Hapus Huruf ganda

Data Tokenizing	Data Hapus Huruf Ganda
['hai', 'kak', 'agar', 'dya', 'bantu', 'lebih', 'lanjut', 'infoin', 'atas', 'nama', 'nama', 'nomor', 'layanan', 'dan', 'alamat', 'lewat', 'dm', 'ya', 'kebijakan', 'privasi', 'layanan', 'indihome', 'bisa', 'di', 'cek', 'melalui', 'makasih', 'dya']	['hai', 'kak', 'agar', 'dya', 'bantu', 'lebih', 'lanjut', 'infoin', 'atas', 'nama', 'nomor', 'layanan', 'dan', 'alamat', 'lewat', 'dm', 'ya', 'kebijakan', 'privasi', 'layanan', 'indihome', 'bisa', 'di', 'cek', 'melalui', 'makasih', 'dya']
['indihome', 'kenapa', 'ya', 'modemnya', 'padahal', 'indikator', 'di', 'lampunya', 'normal2', 'aja', 'tapi', 'waktu', 'mau', 'connect', 'gak', 'bisa', 'dan', 'jaringannya', 'nggak', 'ada', 'di', 'hp']	['indihome', 'kenapa', 'ya', 'modemnya', 'padahal', 'indikator', 'di', 'lampunya', 'normal2', 'aja', 'tapi', 'waktu', 'mau', 'connect', 'gak', 'bisa', 'dan', 'jaringannya', 'nggak', 'ada', 'di', 'hp']
['sekarang', 'mau', 'telepon', 'keluar', 'malah', 'responnya', 'diblokir', 'karena', 'belum', 'bayar', 'gue', 'udah', 'bayar', 'tagihan', 'dari', 'dari', 'kapan', 'tau', 'buka', 'isolir', 'teleponnya', 'sekarang']	['sekarang', 'mau', 'telepon', 'keluar', 'malah', 'responya', 'diblokir', 'karena', 'belum', 'bayar', 'gue', 'udah', 'bayar', 'tagihan', 'dari', 'kapan', 'tau', 'buka', 'isolir', 'teleponya', 'sekarang']
['kak', 'deddy', 'maaf', 'ya', 'bikin', 'gak', 'nyaman', 'agar', 'dya', 'bantu', 'lebih', 'lanjut', 'infoin', 'atas', 'nama', 'nomor', 'layanan', 'dan', 'alamat', 'lewat', 'dm', 'ya', 'kebijakan', 'privasi', 'layanan', 'indihome', 'bisa', 'di', 'cek', 'melalui', 'makasih', 'dya']	['kak', 'deddy', 'maf', 'ya', 'bikin', 'gak', 'nyaman', 'agar', 'dya', 'bantu', 'lebih', 'lanjut', 'infoin', 'atas', 'nama', 'nomor', 'layanan', 'dan', 'alamat', 'lewat', 'dm', 'ya', 'kebijakan', 'privasi', 'layanan', 'indihome', 'bisa', 'di', 'cek', 'melalui', 'makasih', 'dya']
['indihome', 'gangguan', 'ya', 'min']	['indihome', 'gangguan', 'ya', 'min']

Langkah berikutnya melibatkan proses normalisasi untuk mengubah dari kata non formal menjadi formal. Tahap berikutnya, dilakukan proses *stopword* untuk menghapus kata yang tidak bermakna. Hasil Normalisasi dan *stopword* dapat terangkum pada tabel 10.

Tabel 10. Hasil Normalisasi dan *stopword*

Data Normalisasi	Data Stopword
['hai', 'kakak', 'supaya', 'dya', 'tolong', 'lebih', 'lanjut', 'memberikan', 'informasi', 'atas', 'nama', 'nomor', 'layanan', 'dan', 'alamat', 'melalui', 'dm', 'ya', 'kebijakan', 'privasi', 'layanan', 'indihome', 'bisa', 'di', 'cek', 'melalui', 'terima', 'kasih', 'dya']	['hai', 'kakak', 'supaya', 'dya', 'tolong', 'lanjut', 'memberikan', 'informasi', 'atas', 'nama', 'nomor', 'layanan', 'alamat', 'melalui', 'dm', 'kebijakan', 'privasi', 'layanan', 'indihome', 'bisa', 'cek', 'melalui', 'terima', 'kasih', 'dya'] Catatan : data yang dihapus adalah 'lebih', 'dan', 'ya', 'di'
['indihome', 'kenapa', 'ya', 'modemnya', 'padahal', 'indikator', 'di', 'lampunya', 'biasa-biasa', 'saja', 'tapi', 'waktu', 'mau', 'connect', 'tidak', 'bisa', 'dan', 'jaringannya', 'tidak', 'ada', 'di', 'handphone']	['indihome', 'kenapa', 'modemnya', 'padahal', 'indikator', 'lampunya', 'biasa-biasa', 'saja', 'waktu', 'mau', 'connect', 'tidak', 'bisa', 'jaringannya', 'tidak', 'ada', 'handphone'] Catatan : data yang dihapus adalah 'ya', 'di', 'tapi', 'dan', 'di'
['sekarang', 'mau', 'telepon', 'keluar', 'malah', 'responnya', 'diblokir', 'karena', 'belum', 'bayar', 'saya', 'sudah', 'bayar', 'tagihan', 'dari', 'kapan', 'tabu', 'buka', 'isolasi', 'teleponnya', 'sekarang']	['sekarang', 'mau', 'telepon', 'keluar', 'malah', 'responnya', 'diblokir', 'belum', 'bayar', 'saya', 'sudah', 'bayar', 'tagihan', 'kapan', 'tahu', 'buka', 'isolasi', 'teleponnya', 'sekarang'] Catatan : data yang dihapus adalah 'karena', 'dari'
['kakak', 'deddy', 'maaf', 'ya', 'bikin', 'tidak', 'nyaman', 'agar', 'dya', 'bantu', 'lebih', 'lanjut', 'memberikan', 'informasi', 'atas', 'nama', 'nomor', 'layanan', 'dan', 'alamat', 'lewat', 'dm', 'ya', 'kebijakan', 'privasi', 'layanan', 'indihome', 'bisa', 'di', 'cek', 'melalui', 'terima', 'kasih', 'dya']	['kakak', 'deddy', 'maaf', 'bikin', 'tidak', 'nyaman', 'agar', 'dya', 'bantu', 'lanjut', 'memberikan', 'informasi', 'atas', 'nama', 'nomor', 'layanan', 'alamat', 'lewat', 'dm', 'kebijakan', 'privasi', 'layanan', 'indihome', 'bisa', 'cek', 'melalui', 'terima', 'kasih', 'dya'] Catatan : data yang dihapus adalah 'lebih', 'ya', 'dan', 'di'
['indihome', 'gangguan', 'ya', 'admin']	['indihome', 'gangguan', 'admin'] Catatan : data yang dihapus adalah 'ya',

Langkah berikutnya melibatkan proses *stemming* untuk mengembalikan kata-kata ke bentuk dasarnya. Dalam penelitian ini, proses *stemming* dilakukan menggunakan *library python* yang disebut Sastrawi. Hasil pada Proses *stemming* dapat terangkum pada Tabel 11.

Tabel 11. Hasil *Stemming*

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
nomor	nomor
memberikan	beri
layanan	layan
kebijakan	bijak
alamat	alamat
indihome	indihome
privasi	privasi

IV. Kesimpulan dan saran

Kesimpulan pada penelitian ini, Dengan melibatkan langkah-langkah dalam preprocessing, dataset telah dipersiapkan untuk tahapan analisis sentimen lebih lanjut. Data yang telah dibersihkan dan dinormalisasi dapat memberikan hasil yang lebih akurat dan relevan dalam mengevaluasi sentimen pengguna terhadap layanan IndiHome.

Daftar Pustaka

- [1] A. Tobing, E. Simorangkir, M. Situmorang, and P. Y. Purba, "The Influence of Service Quality, Pricing, and Digital Marketing on IndiHome Fiber Customer Loyalty," *International Journal of Social Science and Business*, vol. 5, no. 2, pp. 167–173, 2021, doi: 10.23887/ijssb.v5i2.35523.
- [2] C. Septianarditya and M. Nasir, "Analisis Kepuasan Pelanggan IndiHome berdasarkan Kualitas Pelayanan, Harga, dan Promosi," *Eksos*, vol. 18, no. 1, pp. 71–79, Jun. 2022, doi: 10.31573/eksos.v18i1.440.
- [3] J. Khatib Sulaiman, T. Danirmala, Y. Sulisty Nugroho, and U. Muhammdiyah Surakarta, "Analisis Sentimen Terhadap Topik Kenaikan Harga Bahan Bakar Minyak (BBM) pada Media Sosial Twitter," *Indonesian Journal of Computer Science*.
- [4] R. Kurnia Pratama and P. W. Setyaningsih, "Analisis Komentar Pada Twitter Terhadap Lapangan Kerja Dengan Metode Naïve Bayes," *Jurnal Information System & Artificial Intelligence*, vol. 3, no. 2, pp. 217–227, May 2023.

- [5] R. Yanuarti, "Analisis Media Sosial Twitter Terhadap Topik Vaksinasi Covid-19," *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*, vol. 6, no. 2, 2021, doi: 10.32528/justindo.v6i2.5503.
- [6] R. Rismaya and A. N. Sofyan, "Interjeksi Dalam Komentar Terhadap Cuitan Akun Twitter @ASKNONYM: KAJIAN MORFOLOGI," *MABASAN*, vol. 14, no. 2, pp. 181–194, Dec. 2020, doi: 10.26499/mab.v14i2.354.
- [7] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/mib.v5i2.2835.
- [8] M. Nur, "Perbandingan Levenshtein Distance Dan Jaro-Winkler Distance Untuk Koreksi Kata Dalam Preprocessing Analisis Sentimen Pengguna Twitter," *Jurnal Fokus Elektroda: Energi Listrik, Telekomunikasi, Komputer, Elektronika dan Kendali*, vol. 06, no. 02, pp. 88–93, 2021.
- [9] S. Suparyati, A. Sentimen, P. Tripadvisor, and A. Fathurahman, "Analisis Sentimen Dengan Klasifikasi Naïve Bayes pada Review Hotel Tripadvisor," *Jurnal Ilmiah Informatika (JIF)*, vol. 10, no. 01, pp. 20–24, Mar. 2022.
- [10] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 3, 2020, doi: 10.30865/mib.v4i3.2181.
- [11] T. Krisdiyanto, E. Maricha, and O. Nurharyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers," *Jurnal CoreIT*, vol. 7, no. 1, 2021.
- [12] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, 2021, doi: 10.34148/teknika.v10i1.311.
- [13] I. As'ad, M. A. Asis, H. M. Pakka, R. Mursalim, and Y. B. M. Noor, "K-Nearest Neighbors Analysis for Public Sentiment towards Implementation of Booster Vaccines in Indonesia," *ILKOM Jurnal Ilmiah*, vol. 15, no. 2, 2023, doi: 10.33096/ilkom.v15i2.1561.365-372.
- [14] Zaenal, Y. Salim, and L. Budi Ilmawan, "Analisis Sentimen terhadap Komentar Negatif di Media Sosial Facebook dengan Metode Klasifikasi Naïve Bayes," *Buletin Sistem Informasi dan Teknologi Islam*, vol. 1, no. 4, pp. 259–265, 2020.
- [15] A. E. Budiman and A. Widjaja, "Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 3, 2020, doi: 10.28932/jutisi.v6i3.2892.
- [16] Y. Salim and L. Budi Ilmawan, "Analisis Sentimen terhadap Komentar Negatif di Media Sosial Facebook dengan Metode Klasifikasi Naïve Bayes," *Buletin Sistem Informasi dan Teknologi Islam*, vol. 1, no. 4, pp. 259–265, 2020.