

Analisis Pre-processing Sentimen Terhadap Komentar Layanan Indihome Pada Twitter

Achmad Novanto^a, Dolly Indra^b, Wistiani Astuti^c

Universitas Muslim Indonesia, Makassar, Indonesia

^a13020180039@student.umi.ac.id; ^bdolly.indra@umi.ac.id; ^cwistiani.astuti@umi.ac.id

Received: xx xx xxxx | Revised: xx xx xxxx | Accepted: xx xx xxxx | Published: xx xx xxxx

Abstrak

Dalam era globalisasi yang terus berkembang, peran teknologi informasi menjadi krusial dalam mengubah cara manusia berinteraksi dan mengakses informasi. 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 tweet memiliki tantangan tersendiri, terutama karena penggunaan kata-kata non-baku dan bahasa informal. Oleh karena itu, pentingnya preprocessing 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 tweet. Hasil penelitian diharapkan dapat memberikan wawasan yang lebih akurat mengenai respons pengguna terhadap layanan IndiHome. Melalui langkah-langkah preprocessing 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.

Kata kunci : .analisis sentimen, indihome, twitter, *preprocessing*

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. Layanan menggunakan teknologi informasi dan komunikasi di seluruh dunia bisa membuat banyak tugas menjadi lebih mudah dan cepat daripada metode manual[2]. Dan saat ini, jaringan IndiHome sudah tersebar di seluruh wilayah Indonesia, dan terus berinovasi untuk memenuhi kebutuhan internet yang lebih baik bagi Masyarakat[3].

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[4], [5].

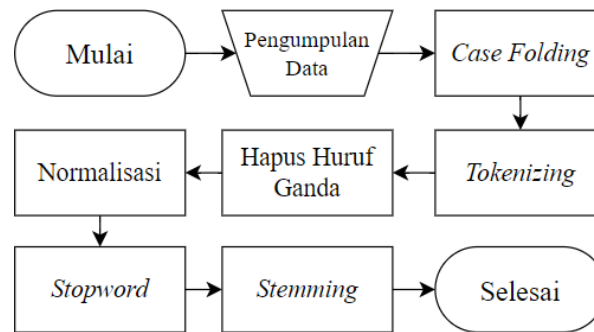
Twitter adalah platform media sosial yang sangat populer di internet. Dengan menggunakan twitter, pengguna dapat berbagi pendapat, mengikuti acara, dan berinteraksi dengan pengguna lainnya. Pertumbuhan penggunaan media sosial membuka peluang baru untuk menganalisis berbagai aspek dan pola dalam komunikasi. Contohnya, data media sosial dapat dianalisis untuk mendapatkan wawasan tentang isu-isu terkini tren, tokoh

berpengaruh, dan jenis informasi lainnya[6]. 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[7].

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[8]. Oleh karena itu, perlu dilakukan *preprocessing* terhadap data *tweet* sebagai tahapan awal dalam analisis sentimen sehingga tepat menghasilkan bentuk data dan meningkatkan keberhasilan klasifikasi[9]. 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[10]. Penelitian ini bertujuan untuk mencari tahu pengaruh dari penerapan teknik preprocessing yang digunakan pada penelitian ini.

Metode

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



Gambar 1. Flowchart tahapan penelitian

A. Pengumpulan Data

Data yang digunakan pada penelitian pada *crawling* dengan menggunakan *tweet harvest* dalam bahasa pemrograman *python*[11]. Data yang akan dicrawling berdasarkan Kumpulan *tweet* dengan kata kunci indihome atau indihomecare dan data disimpan pada bentuk file *Comma Separated Values* (CSV). Hasil *crawling* data terlihat pada Gambar 2.

created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id_str	conversation_id_str	username	tweet_url
Fri Dec 29 23:59:49 +0000 2023	174088536461168629	@indihomeid Hal Kik Ager Dya banyu loh banyu	0	0	0	0	in	7608320886116789	174088433228622678	indihomeCare	https://twitter.com/indihomeCare/status/174088536461168629
Fri Dec 29 23:55:43 +0000 2023	174088433228622678	indihome banyu ya madamepa pantiul indihome	0	1	0	0	in	129192762621096512	174088433228622678	reddhyf	https://twitter.com/reddhyf/status/174088433228622678
Fri Dec 29 23:52:19 +0000 2023	1740882973371836284	@indihomeCare @indihome @TIAZ_KI @SARA000000000000	0	1	0	0	in	918708629	1740882973371836284	@naucikgraj	https://twitter.com/naucikgraj/status/1740882973371836284
Fri Dec 29 23:24:43 +0000 2023	1740876537792346780	@indihome Kik Dedyd mual ya banyu gah nyaman	0	0	0	1	in	7608320886116789	1740844136487507954	indihomeCare	https://twitter.com/indihomeCare/status/1740876537792346780
Fri Dec 29 23:21:18 +0000 2023	1740875662965426129	@Thekay Hal Kik Ager Dya banyu loh banyu	0	0	0	0	in	7608320886116789	1740872626379605347	indihomeCare	https://twitter.com/indihomeCare/status/1740875662965426129
Thu Dec 28 13:51:18 +0000 2023	1740367296509532962	indihome apa banyu sh	0	2	0	0	in	166768448173087792	1740367296509532962	pinkindihome	https://twitter.com/pinkindihome/status/1740367296509532962
Thu Dec 28 13:00:58 +0000 2023	174036172462814808	@gprah, indihome	0	2	0	0	in	1374681389184217091	1740365816270887648	WANTERDHI	https://twitter.com/WANTERDHI/status/174036172462814808
Thu Dec 28 12:59:42 +0000 2023	174036091639567928	@vrommonee indihome 🍷	0	2	0	0	in	1576309330542899914	1740360557960388986	rayaBisa	https://twitter.com/rayaBisa/status/174036091639567928
Thu Dec 28 12:56:53 +0000 2023	1740356144897648643	@indihome Care ya hal @ gis hal janyu ngg	0	3	0	0	in	1633443262531886211	1740341086378469321	ChackK2	https://twitter.com/ChackK2/status/1740356144897648643
Thu Dec 28 12:54:36 +0000 2023	174035569727817223	@indihomeCare mpa kiki nani nani mpa mpa kiki kiki	0	1	0	0	in	17248846622318861048	1740343387728048218	crackabackbar	https://twitter.com/crackabackbar/status/174035569727817223

Gambar 2. Hasil *crawling* data

B. Case Folding

Case Folding adalah suatu proses untuk mengubah semua huruf dalam suatu teks menjadi huruf kecil secara keseluruhan, atau disebut juga *lowercase*[12]. Misalnya, pada kalimat 'Internetnya sangat Cepat Baaangeeet' akan diperkecil hufurnya menjadi 'internetnya sangat cepat baaangeeet'. Adapun contoh dapat dilihat pada Tabel 1.

Tabel 1. Contoh *Case Folding*

Sebelum Case Folding	Setelah Case Folding
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

C. Tokenizing

Tokenizing adalah langkah di mana sekumpulan kata yang membentuk sebuah kalimat akan dipecah menjadi kata-kata tunggal atau token[13]. Menggunakan *Tokenizing* mampu membedakan antara pemisah kata dan non-kata. Misalnya, pada kalimat 'internetnya sangat cepat baaangeeet' dipisahkan menjadi kata 'internetnya', 'sangat', 'cepat' dan 'baaangeeet'. Adapun contoh dapat dilihat pada Tabel 2.

Tabel 2. Contoh *Tokenizing*

Sebelum Tokenizing	Setelah Tokenizing
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']

D. Hapus Huruf Ganda

Hapus huruf Ganda dimana tahapan untuk menngnormalkan huruf berulang. Contoh, pada kata 'baaangeeet' huruf a dan e memiliki huruf ganda prosesnya dihapus huruf ganda menjadi 'banget'. Adapun contoh dapat dilihat 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']

E. 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'. Adapun contoh dapat dilihat 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']

F. 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]. Contoh dapat dilihat 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']

G. Stemming

Stemming merupakan tahap di mana kata-kata dengan imbuhan diubah menjadi bentuk dasarnya[16]. Misal, dalam kata "memulai" diubah menjadi kata "mulai".

Perancangan

Pada penelitian ini dimulai dengan *crawling* data twitter untuk mengumpulkan seluruh dataset dari komentar twitter. Hasil sampel pada dataset terangkum pada Tabel 6.

Tabel 6. 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	@dedydysat 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 7 menampilkan proses hasil dari *case folding*.

Tabel 7. Hasil *case folding*

Data
hai kak agar dya bantu lebih lanjut inforin 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 inforin 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 inforin 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 8.

Tabel 8. Hasil *tokenizing* dan Hapus Huruf ganda

Data Tokenizing	Data Hapus Huruf Ganda
['hai', 'kak', 'agar', 'dya', 'bantu', 'lebih', 'lanjut', 'inforin', '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', 'inforin', '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', 'kapan', 'tau', 'buka', 'isolir', 'teleponnya', 'sekarang']	['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', 'inforin', 'atas', 'nama', 'nomor', 'layanan', 'dan', 'alamat', 'lewat', 'dm', 'ya', 'kebijakan', 'privasi', 'layanan', 'indihome', 'bisa', 'di', 'cek', 'melalui', 'makasih', 'dya']	['kak', 'deddy', 'maaf', 'ya', 'bikin', 'gak', 'nyaman', 'agar', 'dya', 'bantu', 'lebih', 'lanjut', 'inforin', '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 9.

Tabel 9. 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', 'tahu', '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', 'ayanan', 'dan', 'alamat', 'lewat', 'dm', 'ya', 'kebijakan', 'privasi', 'ayanan', 'indihome', 'bisa', 'di', 'cek', 'melalui', 'terima', 'kasih', 'dya']	['kakak', 'deddy', 'maaf', 'bikin', 'tidak', 'nyaman', 'agar', 'dya', 'bantu', 'lanjut', 'memberikan', 'informasi', 'atas', 'nama', 'nomor', 'ayanan', 'alamat', 'lewat', 'dm', 'kebijakan', 'privasi', 'ayanan', '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',

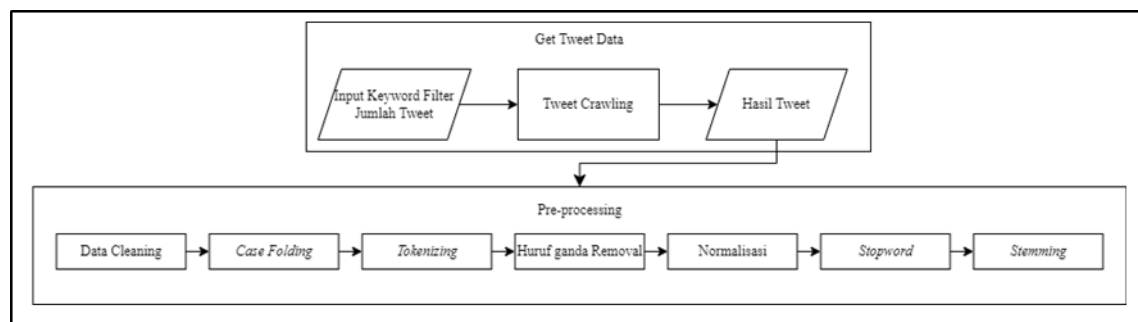
Langkah berikutnya melibatkan proses *stemming* untuk mengembalikan kata-kata ke bentuk dasarnya. Dalam penelitian ini, proses *stemming* dilakukan menggunakan *library python* yang disebut Sastrawi. Hasil pada Proses *stemming* dapat terangkum pada Tabel 10.

Tabel 10. Hasil *Stemming*

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
nomor	nomor
memberikan	beri
ayanan	layan
kebijakan	bijak
alamat	alamat
indihome	indihome
privasi	privasi
bisa	bisa
modemnya	modem
telponnya	telpon
lampunya	lampu
padahal	padahal
...	...
gangguan	ganggu

Pemodelan

Pada penelitian pemodelan yang digunakan adalah pemodelan *flowchart* terlihat pada Gambar 3.



Gambar 3. Pemodelan *Pre-Processing*

Terlihat pada gambar diatas menunjukkan Proses pengambilan data dimana proses mendapatkan data dari database. Proses ini dapat dilakukan mengambil data *tweet* dengan menginput *Keyword Filter* untuk menentukan kata kunci yang digunakan menyaring *tweet*. Selanjutnya, *tweet crawling* akan mengumpulkan data *tweet* yang sesuai dengan kata kunci yang ditentukan seperti kata kunci indihome atau indihomecare. Setelah *crawling* maka akan menjadi dataset pada hasil *tweet* untuk melakukan *pre-processing*.

Preprocessing bertujuan untuk mempersiapkan data yang akan digunakan untuk pelatihan model. Pada

Langkah ini dapat dilakukan pada tahap meliputi:

1. *Data cleaning*, proses untuk membersihkan data yang mengandung *symbol*, numerik, *mention* pada *tweet* dan Uniform Resource Locator (URL).
2. *Case folding*, proses untuk mengubah semua huruf dalam data menjadi huruf kecil.
3. *Tokenizing*, proses untuk memecah data menjadi unit-unit yang lebih kecil, yaitu token.
4. Huruf ganda *removal*, proses untuk menghilangkan huruf ganda yang tidak diperlukan.
5. Normalisasi, proses untuk mengubah data menjadi bentuk yang lebih standar.
6. *Stopword removal*, proses untuk menghilangkan kata-kata yang tidak penting.
7. *Stemming*, proses untuk mengurangi kata menjadi bentuk dasar.

Kesimpulan

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, dan 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, hlm. 167–173, 2021, doi: 10.23887/ijssb.v5i2.35523.
- [2] D. Indra dan J. T. Purba, "Pengukuran Kualitas Pelayanan, Harga dan Customer Value Terhadap Kepuasan Pelanggan Transportasi Online Bagi Kaum Millenial," *Journal of Accounting & Management Innovation*, vol. 3, no. 1, hlm. 14–35, Jan 2019.
- [3] C. Septianarditya dan M. Nasir, "Analisis Kepuasan Pelanggan IndiHome berdasarkan Kualitas Pelayanan, Harga, dan Promosi," *Eksos*, vol. 18, no. 1, hlm. 71–79, Jun 2022, doi: 10.31573/eksos.v18i1.440.
- [4] J. Khatib Sulaiman, T. Danirmala, Y. Sulisty Nugroho, dan U. Muhammdiyah Surakarta, "Analisis Sentimen Terhadap Topik Kenaikan Harga Bahan Bakar Minyak (BBM) pada Media Sosial Twitter," *Indonesian Journal of Computer Science*.
- [5] R. Kurnia Pratama dan P. W. Setyaningsih, "Analisis Komentar Pada Twitter Terhadap Lapangan Kerja Dengan Metode Naïve Bayes," *Jurnal Information System & Artificial Intelligence*, vol. 3, no. 2, hlm. 217–227, Mei 2023.
- [6] 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.
- [7] R. Rismaya dan A. N. Sofyan, "Interjeksi Dalam Komentar Terhadap Cuitan Akun Twitter @ASKNONYM: KAJIAN MORFOLOGI," *MABASAN*, vol. 14, no. 2, hlm. 181–194, Des 2020, doi: 10.26499/mab.v14i2.354.
- [8] S. Khairunnisa, A. Adiwijaya, dan 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, hlm. 406, Apr 2021, doi: 10.30865/mib.v5i2.2835.
- [9] 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, hlm. 88–93, 2021.
- [10] S. Suparyati, A. Sentimen, P. Tripadvisor, dan A. Fathurahman, "Analisis Sentimen Dengan Klasifikasi Naïve Bayes pada Review Hotel Tripadvisor," *Jurnal Ilmiah Informatika (JIF)*, vol. 10, no. 01, hlm. 20–24, Mar 2022.
- [11] R. Tinages, A. Triayudi, dan 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, hlm. 650, Jul 2020, doi: 10.30865/mib.v4i3.2181.
- [12] T. Krisdiyanto, E. Maricha, dan 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.
- [13] H. C. Husada dan 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.
- [14] Zaenal, Y. Salim, dan L. Budi Ilmawan, "Buletin Sistem Informasi dan Teknologi Islam Analisis

- Sentimen terhadap Komentar Negatif di Media Sosial Facebook dengan Metode Klasifikasi Naïve Bayes,” *Buletin Sistem Informasi dan Teknologi Islam*, vol. 1, no. 4, hlm. 259–265, 2020.
- [15] A. E. Budiman dan 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 dan L. Budi Ilmawan, “Buletin Sistem Informasi dan Teknologi Islam Analisis Sentimen terhadap Komentar Negatif di Media Sosial Facebook dengan Metode Klasifikasi Naïve Bayes INFORMASI ARTIKEL ABSTRAK,” vol. 1, no. 4, hlm. 259–265, 2020.