

METODE LEXICON BASED UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP KINERJA ISP (STUDI KASUS : INDIHOME, BIZNET, MYREPUBLIC)

Muhammad Bagas Dikal Putra, Eman Setiawan

Sistem Informasi, Fakultas Ilmu Komputer, Universitas Narotama Surabaya

Jl. Arief Rachman Hakim 51, Sukolilo – Surabaya, Indonesia

bagas.dikal@gmail.com

ABSTRAK

Internet Service Provider (ISP) adalah sebuah perusahaan yang menyediakan layanan koneksi internet atau layanan terkait lainnya kepada pelanggan. ISP telah berkembang seiring dengan perkembangan teknologi. Masih terdapat beberapa masalah yang sering dihadapi oleh penyedia internet seperti Indihome, Biznet, dan MyRepublic. Pemilihan ketiga perusahaan Penyedia Layanan Internet atau *Internet Service Provider* ini didasari karena ketiga ISP yang dipilih memiliki popularitas yang tinggi di Indonesia. Penelitian analisis sentimen telah banyak dilakukan pada media sosial seperti Twitter. Penelitian ini memungkinkan pengguna untuk mengekspresikan opini yang objektif tentang berbagai topik. Dalam penelitian ini, digunakan metode *Lexicon Based* dimana yang merupakan suatu proses yang digunakan untuk memilih kata-kata penting pada dokumen dengan menggunakan kamus yang tersedia. Pada pengaplikasiannya, terdapat dua jenis kamus yang digunakan sebagai *wordlist*, yaitu kamus dengan kumpulan kata bersentimen positif dan negatif. Hasil keseluruhan menunjukkan bahwa sebanyak 80,9% pengguna memberikan sentimen negatif dan 19,1% memberikan sentimen positif terhadap kinerja ketiga penyedia internet tersebut. Terlihat bahwa performa mereka sedang dalam kondisi yang tidak baik. Penyebab pengguna twitter yang memberikan sentimen negatif berasal dari pernyataan para pengguna ISP yang mengalami berbagai permasalahan terhadap ISP itu sendiri.

Kata kunci : Analisis Sentimen, *Lexicon Based*, ISP, Layanan Internet.

1. PENDAHULUAN

Penyedia Layanan Internet (ISP) adalah perusahaan yang memberikan akses internet kepada pelanggan. Awalnya, ISP mengandalkan jaringan telepon, namun kini teknologi serat optik dan nirkabel lebih sering digunakan. Meskipun demikian, ISP seperti Indihome, Biznet, dan MyRepublic masih menghadapi berbagai masalah. Masalah-masalah ini seringkali menjadi topik hangat di media sosial, menghasilkan banyak opini dan sentimen dari pengguna internet.

Pemilihan ketiga perusahaan Penyedia Layanan Internet atau *Internet Service Provider* ini didasari karena ketiga ISP tersebut memiliki popularitas yang tinggi di Indonesia.

Analisis sentimen merupakan teknik yang digunakan untuk mengevaluasi dan mengidentifikasi emosi dan opini, baik positif maupun negatif [1].

Penelitian analisis sentimen telah banyak dilakukan pada media dokumen. Twitter, sebagai salah satu media sosial yang populer, memungkinkan pengguna untuk mengekspresikan opini yang objektif tentang berbagai topik [2].

Penelitian ini, menggunakan metode *Lexicon Based* dimana yang merupakan suatu proses yang digunakan untuk memilih kata-kata penting pada dokumen dengan menggunakan kamus atau *lexicon* yang sudah tersedia. Pada penggunaannya, terdapat dua jenis kamus yang digunakan sebagai *wordlist*, yaitu kamus dengan kumpulan kata bersentimen positif dan kamus dengan kumpulan kata bersentimen negatif [3].

2. TINJAUAN PUSTAKA

Menurut penelitian dari Ahmad Choirun Najib, Akhmad Irsyad, Ghiffari Assamar Qandi, Nur Aini Rakhmawati yang memiliki judul “Perbandingan Metode *Lexicon-Based* Dan SVM Untuk Analisis Sentimen Berbasis Ontologi Pada Kampanye Pilpres Indonesia Tahun 2019 Di Twitter” [4].

Penelitian ini menghasilkan opini positif dan negatif dari pengaruh penggunaan dataset yang mempengaruhi akurasi setiap kata dalam *tweet*. Hal yang perlu dibenahi dari penelitian ini adalah perlu dilakukan proses ekstraksi dan pembuatan dataset secara semi-otomatis yang bertujuan untuk memperbanyak atribut yang belum ditentukan secara manual.

Selanjutnya menurut Nadhif Sanggara Fathullah, Yuita Arum Sari dan Putra Pandu dalam penelitian yang berjudul “Analisis Sentimen Terhadap Rating Dan Ulasan Film Dengan Menggunakan Metode Klasifikasi *Naïve Bayes* Dengan Fitur *Lexicon-Based*” [3].

Penelitian ini menghasilkan nilai yang lebih tinggi daripada hasil klasifikasi karena fitur penghapusan *stopwords* yang menyebabkan adanya perbedaan yang disebabkan oleh fakta bahwa *lexicon* sentimen hanya berisi kumpulan kata-kata dengan sentimen tertentu. Hal yang perlu dibenahi dalam penelitian ini adalah penambahan jumlah data dan jumlah uji untuk mendapatkan hasil yang lebih optimal.

2.1. Analisis Sentimen

Tentang suatu subjek, seperti individu, organisasi, atau produk, dalam kumpulan data yang dikategorikan sebagai sentimen positif maupun sentimen negatif. [5]

Penelitian analisis sentimen telah banyak dilakukan pada media dokumen. Twitter, sebagai salah satu media sosial yang populer, memungkinkan pengguna untuk mengekspresikan opini yang objektif tentang berbagai topik [6].

2.2. Text Mining

Text mining atau penambangan teks adalah proses penambangan yang dilakukan oleh komputer untuk mendapatkan informasi baru atau mengungkapkan sesuatu yang sebelumnya tidak diketahui, serta menemukan kembali informasi yang tersebar di berbagai sumber data teks yang berbeda secara otomatis. [7]

Tahap-tahap *text mining* secara umum adalah *text preprocessing* dan *text transformation* [7]

2.3. Twitter

Twitter adalah platform media sosial yang memungkinkan pengguna untuk berbagi pesan singkat (tweet) dengan maksimal 140 karakter. Fitur-fitur seperti *hashtag* (#) dan mengikuti akun membuat Twitter menjadi alat yang efektif untuk berbagi informasi, berdiskusi, dan berinteraksi dengan orang lain. [8]

2.4. Internet Service Provider (ISP)

Internet Service Provider (ISP) adalah sebuah perusahaan atau organisasi yang menyediakan layanan akses internet dan jasa komunikasi data kepada pengguna internet. ISP bertanggung jawab atas distribusi dan pengaturan lalu lintas data yang melalui jaringan internet.

2.5. Lexicon Based

Metode *Lexicon Based* adalah salah satu cara umum dalam menganalisis sentimen pada media sosial. Metode ini menggunakan kamus sebagai sumber bahasa yang berfungsi untuk mengklasifikasikan sentimen dari setiap opini sehingga kalimat sentimen dapat diklasifikasikan ke dalam kelas negatif dan positif. [9].

2.6. Load Dictionary

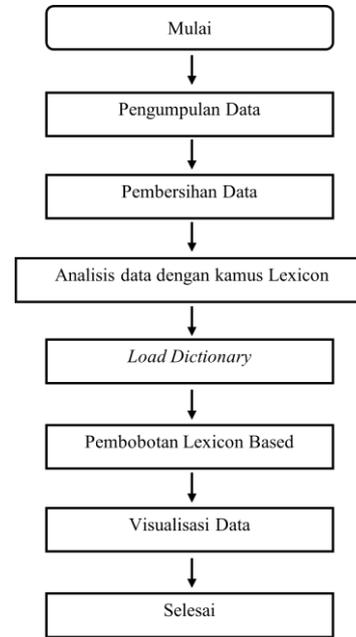
Load Dictionary merupakan salah satu tahap dalam penggunaan metode *Lexicon Based* untuk analisis sentimen. Langkah-langkah yang terlibat dalam proses *load dictionary* terdiri atas : Pemilihan kamus, *Preprocessing*, Format kamus, *Load Dictionary*, dan Mengintegrasikan dengan analisis sentimen.

2.7. Visualisasi Data

Visualisasi data adalah representasi grafis atau visual dari informasi dan data. Tujuan dari data

visualisasi data adalah untuk menyajikan data secara lebih mudah dipahami dan lebih menarik melalui penggunaan elemen visual seperti grafik, diagram, dan sebagainya.

3. METODE PENELITIAN



Gambar 1. Alur penelitian

Metode penelitian ini meliputi pengumpulan data, pembersihan data, dan analisis sentimen menggunakan pendekatan lexicon yang melibatkan pemuatan kamus dan pembobotan fitur, *Load Dictionary*, dan *Pembobotan Lexicon Based*.

3.1. Pengumpulan dan Pembersihan Data

Pada tahap awal pengumpulan data tweet, data diambil dengan menggunakan metode *crawling* dari media sosial Twitter. Proses *crawling* dilakukan dengan bantuan *snscraper* dan *panda* dari *Python Library* dan membutuhkan *API Key* dari Twitter untuk akses sebagai pengembang Twitter pada situs <http://dev.twitter.com> dengan akun Twitter yang pernah dimiliki sebelumnya. Setelah masuk, baru bisa membuat aplikasi baru untuk memperoleh akses pengguna yang digunakan untuk mengakses informasi dari Twitter.

3.2. Analisis Data dengan Lexicon

Proses pemberian label tweet dilakukan dengan menggunakan kamus *lexicon based* dengan melakukan analisis terhadap hasil klasifikasi untuk mengetahui jenis opini apa yang paling sering muncul. Dalam setiap kalimat, dilakukan pencarian nilai sentiment kemudian kata-kata tersebut terklasifikasi dengan rumus (1) dan (2)

$$S_{positive} = \sum_{i \in T}^{n} positive\ score_i \tag{1}$$

$$S_{negative} = \sum_{i \in T}^n negative\ score_i \quad (2)$$

Dalam satu kalimat, bobot dari kalimat diperoleh melalui penjumlahan skor polaritas kata opini positif ($S_{positive}$) dan skor polaritas kata opini negatif ($S_{negative}$) Bobot ini digunakan sebagai acuan untuk proses perbandingan. Dengan demikian, total nilai positif ($S_{positive}$), nilai negatif ($S_{negative}$) dan nilai netral dapat ditentukan dari kata-kata penyusunnya. Persamaan 3 digunakan untuk menentukan orientasi sentimen dengan membandingkan jumlah nilai positif, negatif, dan netral.

$$Sentence_{sentiment} \begin{cases} positive & \text{if } S_{positive} > S_{negative} \\ neutral & \text{if } S_{positive} = S_{negative} \\ negative & \text{if } S_{positive} < S_{negative} \end{cases}$$

Gambar 2. Pemberian label terhadap tweet yang sudah diberi bobot

Dengan demikian, hasil analisis dan pengujian tersebut dapat digunakan untuk memberikan pandangan dan rekomendasi pada pengambilan Keputusan. Setelah data disortir, akan dihasilkan jumlah bobot yang berkaitan dengan data tersebut. Jika total bobotnya positif, data yang positif tersebut dikategorikan sebagai tweet dengan opini yang sesuai.

3.3. Load Dictionary

Load Dictionary merupakan salah satu tahap dalam penggunaan metode *Lexicon Based* untuk analisis sentimen. Berikut adalah Langkah-langkah yang terlibat dalam penelitian ini selama proses *load dictionary* :

- a. Pemilihan kamus: Peneliti menggunakan kamus *WordNet* sebagai kamus sentimen, dimana *WordNet* adalah sebuah lexicon bahasa Inggris yang disusun berdasarkan hierarki semantik. [9]
- b. *Preprocessing*: Pada tahap *preprocessing* ini melibatkan proses menghapus tanda baca, mengubah huruf menjadi huruf kecil, dan menghilangkan *stop words* yang merupakan kata-kata umum yang tidak memberikan banyak informasi sentimen.
- c. Format kamus: Setiap kamus sentimen dapat memiliki format yang berbeda. Beberapa kamus dapat berupa file teks sederhana dengan daftar kata-kata dan polaritas yang sesuai, dalam penelitian ini, peneliti menggunakan file CSV (*Comma-Separated Values*) sebagai media penyimpanan kamus sentimen,
- d. Memuat kamus: Proses ini melibatkan membaca kamus dari *file* yang relevan dan menyimpannya dalam struktur data yang sesuai, seperti daftar kata-kata dengan polaritas yang sesuai atau kamus *hash* yang memetakan kata-kata ke polaritas mereka.
- e. Integrasi dengan analisis sentimen: Proses ini melibatkan pencocokan kata-kata dalam teks dengan kamus sentimen dan menghitung skor

sentimen berdasarkan polaritas yang sesuai. Skor sentimen kemudian dapat digunakan untuk menganalisis dan mengklasifikasikan sentimen teks.

Pada tahap ini, dilakukan pencocokan kata dalam teks dengan kamus kata sentimen dan *emoticon* untuk menentukan kandungan sentimen dari teks tersebut.

3.4. Pembobotan Lexicon Based

Tahap pembobotan *Lexicon Based* dalam analisis sentimen bertujuan untuk memberikan nilai numerik pada kata-kata berdasarkan sentimen yang terkandung di dalamnya. Nilai ini kemudian digunakan untuk menghitung sentimen keseluruhan dari sebuah kalimat. Proses pembobotan ini melibatkan beberapa langkah:

- a. Penentuan skor sentimen: Kata-kata dikategorikan menjadi tiga: positif, negatif, atau netral.
- b. Penentuan skala bobot: Skala yang digunakan biasanya berkisar antara -1 (sangat negatif) hingga +1 (sangat positif).
- c. Pembobotan kata:
 - Penentuan polaritas: Setiap kata dibandingkan dengan kamus sentimen untuk menentukan apakah kata tersebut berpolaritas positif atau negatif.
 - Penanganan negasi: Kata-kata negasi seperti "tidak" akan membalikkan polaritas kata yang mengikutinya.
 - Pemberian skor kalimat: Setelah setiap kata diberi skor, skor-skor ini kemudian digabungkan untuk menghitung skor sentimen keseluruhan dari sebuah kalimat.

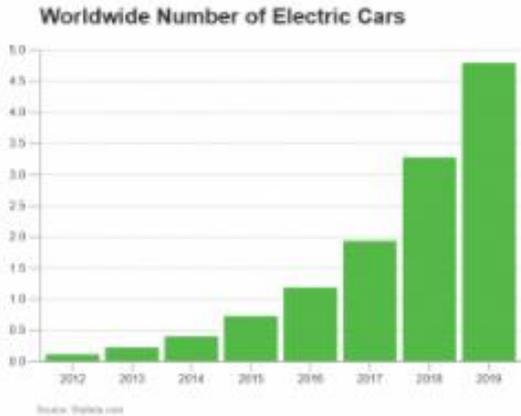
Tujuan utama dari tahap ini adalah untuk mengubah teks yang bersifat kualitatif (positif, negatif, netral) menjadi data kuantitatif yang dapat diproses oleh komputer.

3.5. Visualisasi Data

Tujuan dari penggunaan visualisasi data adalah untuk menyajikan data sentimen pengguna Twitter terhadap kinerja *Internet Service Provider* yang mudah dipahami dan lebih menarik melalui penggunaan elemen visual seperti grafik

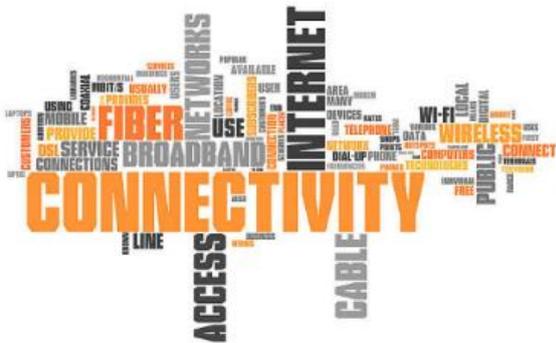
Berikut adalah langkah-langkah yang dilakukan dalam proses visualisasi data:

- a. Pengumpulan Data : Pertama, data yang relevan dikumpulkan dapat berupa angka, teks, atau gambar.
- b. Pemilihan metode visualisasi: Setelah data terkumpul, metode visualisasi yang paling sesuai harus dipilih. Pada penelitian ini menggunakan metode visualisasi berbentuk diagram batang dan WordCloud yang dijelaskan pada gambar berikut Grafik Batang: Digunakan untuk membandingkan nilai atau kategori.



Gambar 3. Contoh Visualisasi data dalam bentuk diagram batang

Word Cloud: Menampilkan kata-kata sentimen yang paling sering muncul dalam teks dengan ukuran yang berbeda sesuai dengan frekuensinya.



Gambar 4. Contoh Visualisasi data dalam bentuk WordCloud

- Pemrosesan data: Data perlu diolah sebelum divisualisasikan, yang melibatkan pemfilteran, penggabungan, atau penghitungan data agar sesuai dengan kebutuhan visualisasi.
- Desain visual: Tahap ini melibatkan pemilihan warna, tata letak, dan gaya visual yang sesuai untuk membuat visualisasi yang jelas dan menarik.
- Pembuatan visualisasi: Tahap ini adalah membuat visualisasi menggunakan alat atau perangkat lunak yang sesuai. Pada penelitian ini akan menggunakan *Python libraries Matplotlib*.
- Interpretasi dan analisis: Setelah visualisasi selesai, data dapat diinterpretasikan dan dianalisis dengan lebih mudah.
- Refining dan iterasi: Menyempurnakan Visualisasi awal berdasarkan umpan balik.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Pada penelitian ini, peneliti menggunakan *tool google colab* dengan Bahasa pemrograman Python, serta menggunakan modul python *snsctraper* dan *pandas* untuk melakukan proses *crawling tweet*. Saat

melakukan proses pengambilan data digunakan *tool Python* ini untuk mengumpulkan data *tweet* dari Twitter. Sebelum melakukan *crawling* data, diperlukan adanya API Twitter yang bertujuan sebagai syarat otentikasi agar mendapatkan hak akses pengambilan data dari Twitter. Kata kunci (*keyword*) juga digunakan dalam proses pengambilan data untuk mencari topik twitter yang sesuai dengan topik penelitian ini yaitu tentang kinerja *Internet Service Provider (ISP)* dari ketiga perusahaan yaitu IndiHome, Biznet, dan Myrepublic. Kode yang digunakan untuk *crawling* data pada gambar berikut:

```

import snsctraper
import snsctraper.modules.twitter as snstwitter
import pandas as pd
import time

tweets_lists = []
maxTweets = 1000
search_key = "Indihome","Biznet","myrepublic"
exclude_username = "@MyRepublicID","@IndiHomeCare","@BiznetHome"
for i,tweet in enumerate(snstwitter.TwitterSearchscraper(" OR ".join(search_key)).get_items()):
    if len(tweets_lists)==maxTweets: break
    if tweet.user.username in exclude_username: continue
    if i%1000 == 0:
        print("Sleep at (%reqs) Request".format(reqs=i))
        time.sleep(10)
    tweets_lists.append([tweet.date, tweet.id, tweet.rawContent, tweet.user.username])
tweets_df = pd.DataFrame(tweets_lists, columns=["Datetime", "Tweet Id", "Text", "Username"])
tweets_df.to_csv('user-tweets_2.csv', sep=',', index=False)
print("done")
    
```

Gambar 5. Kode Crawling data Twitter

Proses pengumpulan data yang dilaksanakan pada tanggal 10 April 2023. Sesudah proses *crawling* yang menghasilkan data sebanyak 5000 *tweet* dengan *keyword* Indihome, dan data sebanyak 1000 data *tweet* dengan *keyword* Biznet dan 1000 data *tweet* dengan *keyword* MyRepublic. Data *tweet* yang didapat masih berupa *raw tweet* yang merupakan data *tweet* yang belum diolah sama sekali. Sehingga diperlukan proses *preprocessing data* untuk membersihkan dan mengubah *raw tweet* menjadi data yang normal dan bisa dianalisa.

Gambar 6. Pengumpulan Data kedalam file .csv

Tabel 1. Atribut pada data tweet

No.	Atribut	Keterangan
1.	"Date"	Waktu pembuatan <i>tweet</i>
2.	"Tweet ID"	ID <i>tweet</i>
3.	"Text"	Isi teks pada <i>tweet</i> yang dibuat
4.	"Username"	Nama Akun pembuat <i>tweet</i>

Sebelum masuk ke dalam tahap Pembersihan data, perlu dilakukan proses pengubahan isi *raw tweet* yang awalnya memiliki banyak atribut hingga menjadi

data yang hanya memiliki atribut teks saja. Data inilah yang akan digunakan sebagai input pada saat tahap Pembersihan Data.

4.2. Tahap Pembersihan Data

Pembersihan data *tweet* dilakukan untuk mengubah data yang awalnya tidak terstruktur menjadi data yang terstruktur sehingga dapat dianalisis dengan mudah.

4.3. Tahap Case Folding

Tahap *Case Folding* berguna untuk mengubah huruf yang ada dalam data set menjadi bentuk huruf kecil atau *lower case*. Berikut adalah kode yang digunakan pada *case folding*:

```
[64] def casefold_text(text):
# Mengganti ke huruf kecil
text=text.lower()

return text

[65] data_tweet['casefold'] = [casefold_text(i) for i in tweet]
tweet = data_tweet['casefold']
```

Gambar 7. Kode Case Folding data Twitter

Pada Gambar 7 merupakan kode yang diberikan berfungsi untuk mengubah semua huruf menjadi huruf kecil (*case folding*). Data yang awalnya disimpan dalam format file *.csv*.

Text	casefold
@TelkomCare gimana cara berenti langganan indih...	gimana cara berenti langganan indihome
@TelkomIndonesia mau langganan indihome	mau langganan indihome
PT Telkom gelar indihome kampung Ramadhan di M...	pt telkom gelar indihome kampung ramadhan di m...
@Rosesvade ANU INDIHOME	anu indihome
@Strategi_Bisnis Pake indihome karena ya cuman...	pake indihome karena ya cuman ini yg ada di at...
#IndiHome #AktifitasTanpaBatas #InternetyaInd...	indihome aktifitastanpabatas internetyaindone...
@TelkomCare mau ganti paket indihome yang lebih...	mau ganti paket indihome yang lebih murah dong
Patut diacungi jempol deh pak Erick Thohir ini...	patut diacungi jempol deh pak erick thohir ini...
Masih ingat julidnya bu Tedjo?Kini Series Tili...	masih ingat julidnya bu tedjokini series tilik...
indihome lag bgttttttt	indihome lag bgttttttt

Gambar 8. Data setelah dilakukan Tahap Case Folding

Pada Gambar 8 menampilkan dari beberapa tweet yang menggunakan huruf kapital mengalami perubahan bentuk huruf menjadi huruf kecil.

4.4. Tahap Punctuation Removal

Tahap *Punctuation Removal* membuang tanda baca yang tidak relevan, seperti titik koma dan tanda seru, untuk menyederhanakan teks dan fokus pada kata-kata yang membawa makna sentimen.. Berikut adalah kode yang digunakan untuk melakukan tahapan *Punctuation Removal*.

```
[71] def removepunc_text(text):
# Hapus tanda baca
text=text.translate(str.maketrans(' ','',string.punctuation))

return text

[72] data_tweet['removepunc'] = [removepunc_text(i) for i in tweet]
tweet = data_tweet['removepunc']
```

Gambar 9. Kode Punctuation Removal data Twitter

Berikut merupakan contoh data yang sudah dilakukan tahap *Punctuation Removal*.

Text	removepunc
wtf with indihome? I should use VPN for max sp...	wtf with indihome should use vpn for max speed...
@MySumatera 1. Bubur Kampiun\2. Lemang Tapai\...	bubur kampiun lemong tapai bika ambon mysunate...
Tanggal 5 sudah lewat aja nih. Kamu jangan lup...	tanggal sudah lewat aja nih kamu jangan lupa c...
@tiramisuhun hai pfp Solon gnteng, how's your ...	hai pfp solon gnteng hows your dayyy
@IndihomeJTD @IndiHome Yuk lah ikutan	yuk lah ikutan
Kan serem ya tiba2 ada yg datang, ngakunya tek...	kan serem ya tiba ada yg datang ngakunya tekni...
@IndiHomeCare @IndiHome wifinya kenapa lagi, b...	wifinya kenapa lagi buat download apa gabisa
koneksi internet super cepat dari IndiHome dan...	koneksi internet super cepat dari indihome dan...

Gambar 10. Data setelah dilakukan Punctuation Removal.

Gambar 10 menunjukkan hasil sampel data yang sudah melalui Tahap *Punctuation Removal* dimana semua simbol dan tanda baca telah dihapus menyisakan hanya karakter huruf dan angka.

4.5. Tahap Stop Word Removal

Tahapan *stop word removal* yaitu tahapan untuk menghilangkan kata-kata umum seperti 'dan', 'yang', dan 'di' (*stop words*) dari teks untuk fokus pada kata-kata yang mengandung sentimen. Daftar *stop words* ini diambil dari repositori *GitHub*. Berikut merupakan beberapa contoh kata yang ada pada kamus *stop word*.

Tabel 2 Contoh kata pada Stop Word

adalah	adapun	agakny	akhirnya
aku	dia	saya	adalah
begitu	dan	dari	kapan
kenapa	siapa	dari	kapan
makanya	mengapa	jadi	sudah

```
from nltk.corpus import stopwords
stopword = stopwords.words('indonesian')
txt_stopword = pd.read_csv('content/drive/MyDrive/TUGAS/stopword.txt', names = ['stopwords'], header = None)

stopword.extend(['kuy', 'yuk', 'hahaha', 'iya', 'pke', 'deh', 'deh', 'kuk', 'dot', 'jg', 'juga', 'yang', 'yg', 'hehe', 'lol', 'lmao', 'an', 'kamu', 'lu', 'mlh', 'dia', 'ng', 'lil', 'the', 'nih', 'gas', 'gaya', 'gas', 'ok', 'kak', 'ha', 'si', 'aja', 'ya', 'lah', 'hah', 'kak', 'wes', 'ada', 'ber', 'langsung', 'knp'])

stopword.extend(txt_stopword['stopwords'][0].split(' '))
stopword = set(stopword)

def stopwords(text):
text = word for word in text if word not in stopword
return text

data_tweet['stopwords'] = data_tweet['token'].apply(stopwords)
data_tweet
```

Gambar 11. Kode Stop Word Removal data Twitter

Gambar 11 merupakan kode untuk melakukan tahapan *Stop word removal*. Berikut contoh data yang telah dilakukan tahapan *stopword removal*.

Text	stopwords
"Pelanggan yang terhormat, Indihome anda akan ...	[pelanggan, terhormat, indihome, blokir, blabl...
tmben cepet indihome	[tmben, cepet, indihome]
@IndiHomeCare Indihome care #KuisIndiHomeCare	[indihome, care, kuisindihomecare]
Mantap jua sudah indihome mun hakun diurus nih...	[mantap, jua, indihome, mun, hakun, diurus, ni...
Yg belum tahu, aku kadih tau nih,\nPerkuat Pos...	[yg, kadih, tau, nih, perkuat, posisi, pasar, ...
...	...
Kalau udah pake Wifi IndiHome di rumah , pulang...	[udah, pake, wifi, indihome, rumah, pulang, ke...
Wah kencang banget internetnya IndiHome. nonto...	[kencang, banget, internetnya, indihome, nonto...

Gambar 12. Contoh Data setelah dilakukan Stop Word Removal.

Gambar 12 merupakan hasil sampel data yang telah melalui tahap *stop word removal*.

4.6. Tahap Stemming

Tahapan stemming bertujuan untuk mengubah kata berimbuhan menjadi bentuk dasarnya. Dalam penelitian ini akan digunakan *library* milik Sastrawi yang merupakan *library* pada bahasa pemrograman python yang dibangun dengan berbasis PHP. Pada penelitian ini tahapan stemming dilakukan bersamaan dengan proses klasifikasi lexicon. Berikut kode yang digunakan pada tahapan *stemming*.

```

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import swifter

factory=StemmerFactory()
stemmer = factory.create_stemmer()

def stemmed_wrapper(term):
    return stemmer.stem(term)

term_dict = {}

tweet = data_tweet['stopwords']
for doc in tweet:
    for term in doc:
        if term not in term_dict:
            term_dict[term]=' '
            print(len(term_dict))
            print("-----")

for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    print(term,":", term_dict[term])

print(len(term_dict))
print("-----")
    
```

Gambar 13. Kode Stemming data Twitter

Berikut adalah contoh hasil ekstraksi data yang telah di *stemming*

gangguannya : ganggu 7848	melengkapi : lengkap 7848
menyimpan : simpan 7848	menguatkan : kuat 7848

Gambar 14. Hasil stemming kata Kembali ke bentuk dasar

Dan berikut adalah hasil data yang telah melalui proses stemming

Text	stemming
Indihome gni amet skrg 😊😞	[indihome, gni, amet, skrg]
Masih ingat julidnya Ibu Tedjo?\n\nKini Series...	[julidnya, tedjo, series, tilik, hadir, cerita...
indihome ngo sih	[indihome, ngo]
benci bgt gw ama indihome hari ini, udh gausah...	[benci, bgt, gw, ama, indihome, udh, gausah, e...
@IndiHome mohon bantuannya untuk sambungan int...	[mohon, bantu, sambung, internet, putus, pagi,...
buset gua banting aja kali ini ya modem myrepu...	[buset, gua, banting, kali, modem, myrepublic]
Lho, logo Twitter-nya kok jadi logo Doge 😊\n\nK...	[lho, logo, twitternya, logo, doge, bos, indiz...

Gambar 15. Hasil data yang telah melalui Tahap Stemming

Hasil data tweet yang telah melalui proses *stemming* seperti yang ditampilkan pada Gambar 15 merupakan hasil dari perubahan kata-kata yang bukan bentuk dasar kemudian dijadikan bentuk dasar misalnya, “Jaringannya” menjadi “Jaringan”.

4.7. Tahap Tokenization

Tahap akhir dari proses *preprocessing* adalah tokenisasi. Pada tahap ini, data teks yang telah dibersihkan akan dipecah menjadi unit-unit terkecil yang disebut token. Pembagian ini dilakukan berdasarkan spasi (space) sebagai pembatas, sehingga setiap kata akan menjadi satu token tersendiri.

```

[137] from nltk.tokenize import RegexpTokenizer
       regexp = RegexpTokenizer(r'\w+|[0-9+|\S+]')
       data_tweet['token'] = data_tweet['clean'].apply(regexp.tokenize)
    
```

Gambar 16. Kode Tokenization data Twitter

Hasil proses *tokenizing* yang dilakukan akan dilakukan perhitungan bobot dengan kamus *Lexicon*. Dan berikut merupakan contoh data yang telah melalui proses tahap *Tokenization*.

Text	token
Masih ingat julidnya bu Tedjo?Kini Series Tili...	[masih, ingat, julidnya, bu, tedjokini, series...
Dengan begini, ke depannya seluruh layanan Ind...	[dengan, begini, ke, depannya, seluruh, layana...
Gengs, sembari ngabuburit bisa nih kita nonton...	[gengs, sembari, ngabuburit, bisa, nih, kita, ...
Masih ingat julidnya Ibu Tedjo?Kini Series Til...	[masih, ingat, julidnya, ibu, tedjokini, serie...
banyak nanya, ga ada solusi. \nteknisi dateng ...	[banyak, nanya, ga, ada, solusi, teknisi, date...
Dari sekadar nonton TV, bisa jadi belajar ba...	[dari, sekadar, nonton, tv, bisa, jadi, belaja...
Paket bundling IndiHome halo! Garcep buruan \n...	[paket, bundling, indihome, halo, garcep, buru...

Gambar 17. Hasil data yang telah melalui Tokenization

4.8. Tahap Klasifikasi dan Pembobotan Data

Proses *preprocessing* data dan *crawling* telah menghasilkan 7000 data bersih yang siap untuk dianalisis sentimennya. Dalam penelitian ini, kami akan menggunakan kamus *InSet Lexicon* yang dirancang khusus untuk menganalisis sentimen pada teks bahasa Indonesia. Kamus ini sebelumnya telah diuji coba oleh Koto dan Rahmaningtyas [10] dalam penelitian mereka terhadap sentimen di Twitter dan terbukti efektif.

Kamus *InSet Lexicon* adalah sebuah daftar kata yang telah dikategorikan berdasarkan sentimen yang ada di dalamnya. Setiap kata dalam kamus ini diberikan skor atau bobot yang menunjukkan seberapa kuat kata tersebut mengekspresikan sentimen positif atau negatif. Serta ditemukan beberapa daftar kata positif dan negatif yang digunakan pada penelitian ini, beberapa diantaranya terdapat pada tabel berikut.

Tabel 3 Jumlah Kata Sentimen Positif dan Negatif pada setiap data

Jenis Sentimen	Jumlah Kata data Indihome	Jumlah Kata data Biznet	Jumlah Kata data MyRepublic
Positif	1083	137	120
Negatif	3917	863	880
Total	5000	1000	1000

Proses pembobotan kata dilakukan dengan cara mencocokkan setiap kata dalam suatu tweet dengan kamus sentimen. Kata-kata yang ditemukan dalam kamus positif diberi nilai +1, sedangkan kata-kata negatif diberi nilai -1. Selanjutnya, dilakukan

penanganan negasi untuk membalikkan polaritas kata yang didahului oleh kata-kata seperti "tidak" atau "bukan". Nilai-nilai dari setiap kata dalam satu tweet kemudian dijumlahkan untuk memperoleh skor sentimen total. Jika skor total positif, maka tweet tersebut dinyatakan memiliki sentimen positif. Sebaliknya, jika skor total negatif, maka tweet tersebut memiliki sentimen negatif. Pada tabel 4 merupakan contoh perhitungan skor polaritas sentimen dengan menggunakan rumus yang ada pada rumus 3.1 dan 3.2.

Tabel 4 Contoh perhitngan skor polaritas sentimen "indihomecare indihome jaringan daerah surabaya error putus sambung koneksi"

Kata sentimen yang terdapat dalam tweet	Error (-1), putus (-1)	
Nilai polaritas sentimen	-2	Negatif
Skor polaritas sentimen	$Sentiment\ Score = \frac{-2}{2} = -1$	Sentimen Negatif

4.9. Hasil Analisis Sentimen

Berdasarkan pengujian yang telah dilakukan pada Penelitian Analisis Sentimen ini, dapat mengasilkan bahwa

- Analisis sentimen dengan menggunakan kata kunci masing-masing layanan penyedia internet. Analisis tersebut menghasilkan sentimen-sentimen yang berbeda dari Jumlah sentimen yang didapatkan setelah menggabung dalam satu data terdapat pada tabel 5 berikut.

Tabel 5 Total Penghitungan Sentimen

Nama Data	Sentimen Positif	Sentimen Negatif
Indihome	1083	3917
Biznet	137	863
MyRepublic	120	880
Total Data	7000	

- Berdasarkan hasil analisis sentimen yang terdapat pada tabel 5 dari sisi jumlah sentimen, jumlah sentimen negatif lebih banyak dibandingkan dengan sentimen positif, karena tidak banyak pengguna twitter yang membuat tweet mengenai Tindakan positif yang dilakukan oleh ketiga layanan penyedia internet tersebut.
- Faktor-faktor yang mempengaruhi dari kinerja ketiga penyedia tersebut diantaranya adalah permasalahan jaringan internet yang tidak stabil, kemudian ada permasalahan mengenai kurangnya jangkauan layanan internet yang disediakan. serta layanan purna jual yang kurang memuaskan.

4.10. Visualisasi Data

Visualisasi data yang digunakan dalam penelitian ini adalah hasil *stopwords* yang ada pada data Twitter pada tanggal 10 April 2023. Hasil visualisasi data yang berbentuk *WordCloud* dapat dilihat pada gambar berikut.

- Naïve Bayes dengan Fitur Lexicon-Based,” 2020. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [4] A. C. Najib, A. Irsyad, G. A. Qandi, and N. A. Rakhmawati, “Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter,” *Fountain of Informatics Journal*, vol. 4, no. 2, pp. 41–48, Nov. 2019, doi: 10.21111/FIJ.V4I2.3573.
- [5] T. Nasukawa and J. Yi, “Sentiment analysis,” in *Proceedings of the 2nd international conference on Knowledge capture*, New York, NY, USA: ACM, Oct. 2003, pp. 70–77. doi: 10.1145/945645.945658.
- [6] L. F. S. Coletta, N. F. F. da Silva, E. R. Hruschka, and E. R. Hruschka, “Combining Classification and Clustering for Tweet Sentiment Analysis,” in *2014 Brazilian Conference on Intelligent Systems*, Sao Paulo, Brazil: IEEE, Oct. 2014, pp. 210–215. doi: 10.1109/BRACIS.2014.46.
- [7] R. Feldman, Ronen, Sanger, and James, *The text mining handbook: Advanced approaches in analyzing unstructured data*. 2007. Accessed: May 10, 2023. [Online]. Available: <https://dl.icdst.org/pdfs/files/25a6d982ee80e1db7a4ebf7eeca4e0ec.pdf>
- [8] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, “Sentiment Analysis of Twitter Data.” [Online]. Available: <http://www.webconfs.com/stop-words.php>
- [9] G. A. Miller, “WORDNET: A LEXICAL DATABASE FOR ENGLISH.” Accessed: May 10, 2023. [Online]. Available: <https://dl.acm.org/doi/pdf/10.1145/219717.219748>
- [10] F. Koto and G. Y. Rahmaningtyas, “Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs,” in *Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017*, Institute of Electrical and Electronics Engineers Inc., Jul. 2017, pp. 391–394. doi: 10.1109/IALP.2017.8300625