

Sentiment Analysis Penyedia layanan Asuransi dari Media Sosial Twitter

Nalar Istiqomah^{1,*}, Fanny Novika²

^{1,2}Aktuaria, Sekolah Tinggi Manajemen Asuransi Trisakti, Jakarta, Kota, Indonesia

Email: ^{1,*} nalar.istiqomah23@gmail.com, ²novikafanny@gmail.com

^{*}) Email Penulis Utama

Abstrak– Tujuan dari penelitian ini adalah menerapkan analisis sentimen untuk mengevaluasi pandangan masyarakat terhadap penyedia layanan asuransi kesehatan, sehingga dapat memahami reputasi penyedia layanan asuransi. Penelitian ini menggunakan algoritme Naïve Bayes untuk mengidentifikasi sentimen pengguna Twitter terkait penyedia layanan asuransi kesehatan dan mengevaluasi akurasi hasilnya. Penelitian ini dilakukan untuk mendapatkan informasi sentimen masyarakat tentang perspektif penyedia layanan asuransi kesehatan melalui media sosial twitter. Pertumbuhan pengguna Twitter yang mencapai lebih dari 500 juta twit setiap hari memberikan potensi *Big Data* untuk mengevaluasi pandangan masyarakat terhadap asuransi kesehatan. Penelitian ini menggunakan metode penelitian deskriptif kualitatif dengan dukungan bahasa pemrograman Python. Populasi penelitian ini mencakup semua twit yang diposting oleh pengguna di Indonesia. Kami menggunakan metode *purposive sampling*, yaitu pemilihan sampel berdasarkan kriteria tertentu yang sesuai dengan tujuan penelitian, seperti twit yang berhubungan dengan penyedia layanan asuransi kesehatan. Data yang digunakan adalah data primer, yaitu twit dari pengguna Twitter di Indonesia yang berkaitan dengan penyedia layanan asuransi kesehatan. Pengumpulan data dilakukan melalui *web scraping* dari aplikasi Tweet Harvest, dilanjutkan dengan proses *labeling*, dan kemudian data dipraproses melalui tahapan pembersihan, tokenisasi, penyaringan, dan *stemming*. Terakhir, algoritme Naïve Bayes digunakan untuk analisis sentimen. Dari proses pengambilan data, kami berhasil mengumpulkan 31.190 data, yang kemudian disaring menjadi 1.483 data yang hanya berupa hasil *review*. Pada tahap pelabelan, 889 twit mendapatkan label positif, sementara 594 twit mendapatkan label negatif. Didapatkan akurasi dari algoritme Naïve Bayes adalah 74.41%. Hasil ringkasan twit menggambarkan pandangan yang beragam terkait BPJS Kesehatan dan asuransi kesehatan swasta. Terdapat pandangan positif terhadap BPJS Kesehatan, termasuk premi yang terjangkau, cakupan penyakit kritis, dan pendaftaran tanpa medical check-up. Namun, ada kritik terhadap prosedur pengobatan yang dianggap rumit, kesulitan dalam menghentikan keanggotaan, dan perdebatan tentang prinsip gotong royong. Di sisi lain, asuransi kesehatan swasta mendapatkan pandangan positif karena prosedur yang lebih sederhana, antrian cepat, dan pilihan produk yang sesuai dengan penghasilan individu. Namun, terdapat juga pandangan negatif, termasuk gangguan *telemarketing*, kasus gagal bayar oleh penyedia asuransi, dan konsumen yang merasa tertipu ketika asuransi pendidikan beralih menjadi asuransi kesehatan tanpa persetujuan mereka.

Kata Kunci: BPJS Kesehatan; Naïve Bayes; Penyedia layanan Asuransi; Sentiment Analysis; Twitter

Abstract– The aim of this research is to apply sentiment analysis to assess the public's opinions of health insurance companies, in order to gain insights into the reputation of these companies. This research employed the Naïve Bayes algorithm to identify sentiments expressed by Twitter users concerning health insurance companies and to evaluate the accuracy of the results. This research was conducted to obtain information on public sentiment regarding the perspective of health insurance service providers through Twitter media. The rapid growth of Twitter users, with over 500 million tweets posted daily, presents an opportunity for utilizing Big Data to assess public sentiment regarding health insurance. This research utilizes a qualitative descriptive research methodology, supported by various tools such as Tweet Harvest, Microsoft Excel, and Python. The research population encompasses all tweets posted by users in Indonesia. We applied a purposive sampling method, which involves selecting samples based on specific criteria aligned with the research objectives. These criteria include tweets related to health insurance companies using specific keywords. The data used consists of primary data, namely tweets from Twitter users in Indonesia related to health insurance companies. Data collection was carried out through web scraping from the Twitter application, followed by a labeling process. Subsequently, the data underwent preprocessing stages, including cleaning, tokenization, filtering, and stemming. Finally, the Naïve Bayes algorithm was applied for sentiment analysis. During the data collection process, we successfully gathered 31,190 data points, which were then filtered to include 1,483 data points representing review outcomes. In the labeling stage, 889 tweets received a positive label, while 594 tweets were assigned a negative label. It was found that the accuracy of the Naïve Bayes algorithm was 74.41%. The summary of tweets illustrates diverse views regarding BPJS Health and private health insurance. There is a positive view of BPJS Health, including affordable premiums, critical illness coverage, and registration without medical check-ups. However, there has been criticism of the treatment procedures as complicated, difficulties in discontinuing membership, and debate about the principle of mutual cooperation. On the other hand, private health insurance is gaining positive views due to simpler procedures, quick queues, and a choice of products that suit individual income. However, there are also negative views, including telemarketing interference, cases of failure to pay by insurance providers, and consumers who feel cheated when education insurance turns into health insurance without their consent.

Keywords: BPJS Health; Naïve Bayes; Insurance Company; Sentiment Analysis; Twitter

1. PENDAHULUAN

Asuransi adalah perjanjian antara penanggung (biasanya Perusahaan asuransi) dan masyarakat sebagai tertanggung (pemegang polis). Masyarakat sebagai pemegang polis berhak mendapatkan proteksi atas penggantian ganti rugi dari penanggung sesuai dengan ketentuan pada polis. Di sisi lain, pemegang polis mempunyai kewajiban untuk melakukan pembayaran premi kepada perusahaan asuransi. Sebaliknya, Perusahaan asuransi berkewajiban untuk membayarkan ganti rugi saat pemegang polis mengalami suatu kemalangan saat pembayaran premi telah diterimanya [1].

Terdapat banyak jenis asuransi, salah satu yang paling banyak digunakan yaitu asuransi kesehatan. Asuransi kesehatan adalah bentuk perlindungan yang menjanjikan kepada pemegang polis untuk mengganti semua pengeluaran medis, seperti biaya perawatan di fasilitas kesehatan, biaya operasi, dan biaya obat-obatan yang diperlukan[2]. Dengan memiliki asuransi kesehatan, pemegang polis atau peserta mendapatkan jaminan bahwa biaya-biaya medis tertentu akan ditanggung oleh penyedia layanan asuransi. Di Indonesia, semakin banyak penduduk yang mendapatkan perlindungan asuransi kesehatan dari tahun ke tahun. Pada tahun 2014, jumlahnya mencapai 133.423.653 orang, sementara pada tahun 2015 meningkat menjadi 156.790.287 orang. Pada tahun 2016, jumlah tertanggung mencapai 171.939.254 orang, dan pada tahun 2017, mencapai 187.982.949 orang. Pada tahun 2018, jumlah tertanggung asuransi kesehatan melonjak menjadi 208.054.199 orang [3]. Tercatat hingga tahun 2018 sudah lebih dari 75% masyarakat Indonesia telah dicakup oleh asuransi kesehatan. Peningkatan ini perlu dianalisis sejalan dengan reputasi asuransi kesehatan di mata masyarakat.

Disamping itu, penting juga untuk mempertimbangkan reputasi penyedia layanan. Reputasi penyedia layanan adalah aset berharga yang memberikan keunggulan dalam persaingan, karena penyedia layanan akan dianggap sebagai entitas yang terpercaya, dapat diandalkan, dan bertanggung jawab terhadap karyawan, pelanggan, pemegang saham, dan pasar keuangan. Reputasi yang baik adalah indikator penting dari kepercayaan dan integritas penyedia layanan [4]. Kepercayaan yang diberikan oleh pemangku kepentingan seperti pelanggan, karyawan, investor, dan masyarakat umum sangat penting untuk pertumbuhan dan kelangsungan hidup penyedia layanan. Reputasi penyedia layanan bukan hanya tentang bagaimana penyedia layanan dinilai oleh masyarakat, tetapi berdampak pada kinerja bisnis dan kesuksesannya. Mempertahankan dan meningkatkan reputasi penyedia layanan harus menjadi prioritas utama dalam strategi bisnis jangka panjang [5]. Penyedia layanan yang dihormati oleh masyarakat umum cenderung mendapatkan dukungan dalam berbagai aspek, termasuk dukungan dalam proyek lokal dan dukungan dalam upaya berkelanjutan.

Banyak cara bisa dilakukan untuk mengetahui reputasi penyedia layanan, salah satunya dengan melihat *review* nasabah atau pengguna asuransi di media sosial. Twitter adalah salah satu *platform* media sosial yang paling populer di Indonesia. Twitter adalah layanan yang memungkinkan teman, keluarga, dan rekan kerja untuk berkomunikasi dan menjaga koneksi melalui pertukaran pesan yang cepat dan mempunyai intensitas yang sering digunakan[6]. Pengguna dapat memposting *tweet*, yang dapat berisi foto, video, tautan, dan teks. Adapun *tweet* adalah setiap pesan yang diposting ke Twitter yang dapat berisi *Big data* dalam bentuk dokumen foto, video, tautan, serta teks [7]. *Big Data* perlu diinterpretasikan sebagai kumpulan informasi besar yang ditandai oleh besarnya volume, keragaman, dan tingkat kompleksitas. Menerapkan data secara luas dalam operasional penyedia layanan asuransi dapat meningkatkan nilai yang diberikan oleh asuransi pada berbagai tingkatan, dan dengan demikian memberikan keunggulan kompetitif. *Big data* memungkinkan peningkatan dalam penyesuaian penawaran penyedia layanan asuransi kepada klien[8], dalam hal ini seluruh *tweet* yang berkaitan dengan asuransi kesehatan agar dapat mengetahui reputasi asuransi kesehatan di Indonesia.

Penelitian ini bermaksud untuk mengetahui pendapat masyarakat terkait asuransi kesehatan di Indonesia melalui pelaporan di Twitter. Memberikan saran atau rekomendasi melalui *platform* media sosial telah menjadi peran yang sangat signifikan dalam beragam sektor industri dimana interaksi dan respon dari pelanggan memiliki peran yang krusial dalam kesuksesan bisnis[9]. Twitter telah menjadi aplikasi media sosial ketujuh yang paling banyak digunakan di dunia [10]. Pengguna Twitter di Indonesia mencapai 18,45 juta pada tahun 2022 [11]. Setiap harinya rata-rata ada 500 juta *tweet* diposting setiap hari[12]. Untuk mendapatkan informasi mengenai penyedia layanan asuransi dari sekian banyak *tweet* itu, dibutuhkan salah satu teknik *text mining* yaitu *sentiment analysis*. *Sentiment analysis*, juga dikenal sebagai penambangan opini, merupakan subdisiplin dari *text mining* yang berfokus pada analisis pandangan, sentimen, penilaian, evaluasi, sikap, dan emosi yang diekspresikan oleh seseorang pembicara atau penulis terhadap suatu topik, produk, layanan, organisasi, individu, atau kegiatan tertentu [13]. Tugas utama dari *sentiment analysis* adalah mengelompokkan teks, kemudian menentukan pendapat yang dikemukakan dalam teks tersebut bersifat positif atau negatif. *Sentiment analysis* juga bisa mengidentifikasi perasaan emosional sedih, gembira, atau marah. Dalam penelitian ini, *sentiment analysis* akan digunakan pada data Twitter untuk melihat reputasi penyedia layanan asuransi kesehatan.

Penelitian tentang *sentiment analysis* telah beberapa kali dilakukan dalam berbagai bidang untuk melihat pandangan masyarakat di media sosial. Hasil *review platform* belanja online dapat dirangkum masukan dari para konsumen menggunakan *sentiment analysis*[14]. Penelitian lain juga dilakukan untuk mengetahui pandangan masyarakat tentang penggunaan vaksin Covid-19 melalui pendekatan Naïve Bayes [15]. Perbedaan dari penelitian

sebelumnya dibandingkan penelitian ini adalah pada objek penelitian, di mana penelitian ini secara spesifik ingin mengetahui pandangan masyarakat tentang asuransi kesehatan.

Sehingga tujuan dari penelitian ini adalah:

1. Menerapkan *sentiment analysis* untuk mengetahui sentimen/pendapat masyarakat pada penyedia layanan asuransi kesehatan, sehingga bisa diketahui reputasi penyedia layanan asuransi.
2. Menggunakan algoritme Naïve Bayes untuk mengidentifikasi sentimen *user* Twitter terhadap penyedia layanan asuransi kesehatan serta mengevaluasi akurasi hasilnya.

2. METODE PENELITIAN

2.1 Metode Penelitian

Penelitian ini menganalisis sentimen *user* Twitter terhadap penyedia layanan asuransi menggunakan algoritme Naïve Bayes. Metode yang digunakan dalam penelitian ini adalah jenis metode penelitian deskriptif kualitatif [16]. Penelitian deskriptif adalah jenis penelitian yang bertujuan untuk secara sistematis dan akurat menggambarkan gejala, fakta, atau kejadian yang terkait dengan sifat populasi atau wilayah tertentu. Dalam penelitian deskriptif, tidak ada upaya untuk menentukan hubungan antara variabel atau menguji hipotesis. Tujuannya adalah menggambarkan situasi, peristiwa, atau profil individu atau masyarakat seperti mengambil foto keadaan tersebut [17]. Penelitian ini bersifat menjabarkan suatu fenomena yaitu sentimen *user* Twitter terhadap penyedia layanan asuransi kesehatan. Penelitian ini dilakukan dengan cara mengumpulkan dan menganalisis data berdasarkan pada karakteristik tertentu yaitu *user* Indonesia dan *tweet* yang membahas penyedia layanan asuransi kesehatan. Dalam melakukan penelitian ini, peneliti dibantu oleh beberapa *tools* aplikasi sebagai alat bantu. Adapun *tools* tersebut adalah:

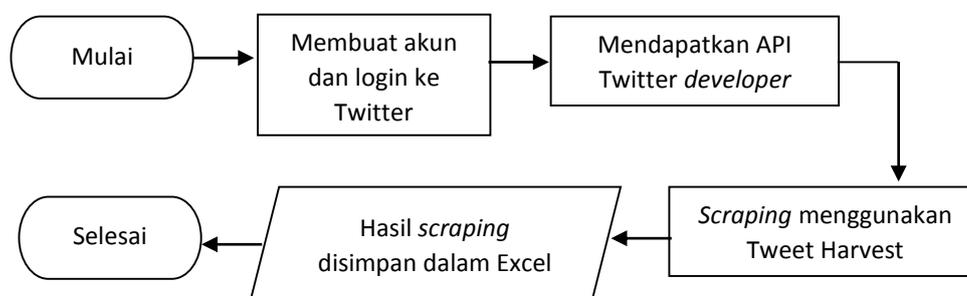
1. Tweet Harvest untuk *crawling* atau *scraping* data Twitter
2. Microsoft Excel untuk menyimpan data hasil *scraping* Twitter
3. Bahasa pemrograman Python untuk pengolahan data

2.2 Populasi dan Sampel

Populasi dalam penelitian ini merujuk pada seluruh *tweet* yang diposting oleh pengguna dari Indonesia. Karena jumlah *tweet* yang berasal dari Indonesia di *platform* Twitter sangat besar, dan tidak semuanya relevan dengan tujuan penelitian, maka diterapkan metode pemilihan sampel. Dalam penelitian ini, digunakan metode *purposive sampling*, dimana sampel dipilih berdasarkan kriteria khusus yang relevan dengan tujuan penelitian. Kriteria sampel yang digunakan mencakup *tweet* yang berasal dari Indonesia dan berhubungan dengan penyedia layanan asuransi kesehatan, seperti *tweet* yang mencakup kata kunci tertentu seperti penyedia layanan asuransi kesehatan, BPJS kesehatan, layanan asuransi kesehatan, pengecualian asuransi kesehatan, dan pelayanan asuransi kesehatan.

2.3 Teknik Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data primer yaitu data *tweet user* Twitter di Indonesia tentang penyedia layanan asuransi kesehatan. Adapun pengumpulan data, dilakukan dengan cara *scraping* menggunakan aplikasi Tweet Harvest. Kegiatan pengolahan data, atau dengan kata lain, ekstraksi data dari internet atau *platform* media sosial, bertujuan untuk mengungkapkan pola-pola yang tersembunyi di dalamnya inilah yang disebut *scraping* [18]. Data yang dikumpulkan merupakan seluruh *tweet* mulai dari tahun Januari 2020 hingga September 2023. Tahapan pengumpulan data dijelaskan pada Gambar 1.



Gambar 1. Tahap pengumpulan Data

2.4 Teknik Analisis Data

Sebelum melakukan analisis data, dilakukan praproses data terlebih dahulu. Tahapan ini bertujuan untuk membersihkan data dan menyesuaikan format data sehingga data bisa dimasukkan kedalam algoritma Naïve Bayes. Pada tahap praproses, terdapat beberapa langkah yang harus dilakukan, yaitu:

1. *Cleaning*, pada tahap ini dilakukan *case folding* atau merubah teks menjadi huruf kecil. Selain itu, dilakukan pula hal-hal seperti menghapus karakter selain huruf, menghapus *username* pengguna atau *mentions* (@), menghapus *hashtag* (#), menghapus URL atau *link* dan menghapus *emoticon*.
2. *Tokenizing*, pada tahapan *tokenizing*, setiap kata akan dipisahkan berdasarkan spasi yang ditemukan.
3. *Filtering*, pada tahap ini dilakukan pembuangan kata-kata tidak penting dari hasil *tokenizing*. Selain itu, dilakukan juga penghapusan tanda baca dan *stopword*. *Stopword* diproses pada sebuah kalimat jika mengandung kata-kata yang sering keluar dan dianggap tidak penting seperti kata penghubung (yang, di, ke, dari, dan).
4. *Stemming*, pada tahapan *stemming*, dilakukan pengubahan kata berimbuhan menjadi kata dasar, misal dari ditulis menjadi tulis.

Setelah dilakukan praproses, dilakukan proses *labeling*. Pada tahap ini dilakukan pemberian label pada tiap data twit berdasarkan isi twit -nya. Label berupa *tag* positif dan negatif. Contoh *labeling* dijelaskan pada Tabel 1.

Tabel 1. Contoh *Labeling*

twit	Label
sampe detik ini gak akan pernah maafin asuransi bumi putera karena gak cairin duit hak orang2 yang ikut asuransinya. kepada petinggi bumi putera dan seluruh orang yang kecipratan duitnya, i wish you a very painful death, i wish you a very rot in hell.	Negatif
Selalu amaze gue sm kantor yg ngasih asuransi tp gaada limitnya, perawatan gigi berjutjut tp pulang2 gak bayar samsek tuh bener2 benefit+++++++++. dokter senang bisa ngasih perawatan maksimal, pasien senang krn semua dicover asuransi	Positif

Labeling dilakukan secara otomatis dengan menggunakan program Python. Kemudian dicek kembali secara manual untuk memastikan label sudah benar.

Data yang telah dipraproses dan diberi label, dapat dipakai untuk klasifikasi menggunakan Naïve Bayes. Berikut adalah tahapan yang dilakukan:

1. Pembagian data menjadi data latih dan data uji. Data latih sebanyak 80% akan digunakan untuk membuat model klasifikasi dengan Naïve Bayes. Sedangkan data uji sebanyak 20% digunakan untuk menguji model klasifikasi yang dihasilkan apakah sudah cukup akurat atau belum.
2. Naïve Bayes adalah teknik klasifikasi yang menggunakan Teorema Bayes. Teknik klasifikasi ini akan memprediksi kelas data dengan menghitung peluangnya. Berikut adalah rumus yang digunakan:

$$P((A|B)) = \frac{P(B|A) * P(A)}{P(B)}$$

Keterangan:

$P(A | B)$: Probabilitas A terjadi dengan bukti bahwa B telah terjadi (probabilitas superior)

$P(B | A)$: Probabilitas B terjadi dengan bukti bahwa A telah terjadi

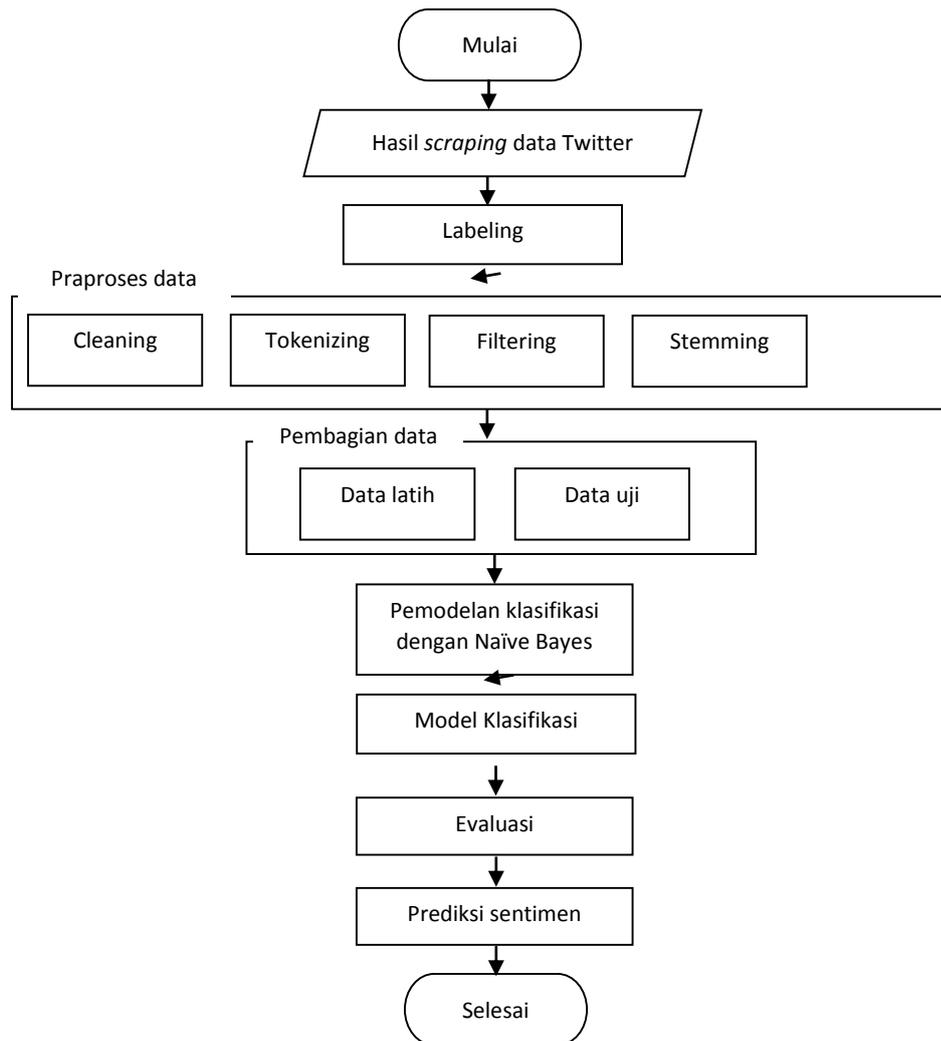
$P(A)$: Peluang terjadinya A

$P(B)$: Peluang terjadinya B

Intinya, Teorema Bayes menemukan probabilitas ketika kita mengetahui probabilitas tertentu lainnya. Pada pemograman Python, kita tidak perlu menghitung peluangnya secara manual. Cukup menggunakan fungsi pada Python yaitu `MultinomialNB()`.

3. Setelah model klasifikasi selesai, model bisa di gunakan untuk memprediksi sentimen data baru.

Tahapan analisis data terdapat pada Gambar 2.



Gambar 2. Tahap Analisis Data

3. HASIL DAN PEMBAHASAN

3.1 Pembahasan Hasil Data *Scrapping*

Data yang dikumpulkan berasal dari twit masyarakat yang mengandung kata kunci asuransi kesehatan, BPJS dan beberapa penyedia layanan asuransi kesehatan seperti AIA, FWD, Generalie, Mandiri in Health dan lain sebagainya. Sebelum melakukan proses *scrapping* data Twitter, kita harus mempunyai akun Twitter terlebih dahulu. Hal ini dilakukan agar kita mendapatkan token atau *auth_token* sebagai penanda bahwa kita memiliki akses ke halaman pencarian Twitter. Sejak bulan Juli 2023, Twitter telah membatasi twit yang dapat dibaca atau di-*scarpping* per harinya, yaitu:

1. Akun *verified* (centang biru) dapat membaca hingga 6000 twit per hari.
2. Akun tidak *verified* (akun biasa) hanya dapat membaca 600 twit per hari
3. Akun baru hanya bisa membaca 300 twit per hari.

Dikarenakan adanya pembatasan, proses *scrapping* dilakukan setiap harinya dengan melibatkan beberapa akun Twitter secara bersamaan. Penelitian ini menggunakan Tweet Harvest, sebuah *software* yang menggunakan bahasa pemrograman Python. Dalam proses ini, kami menggunakan *auth_token* Twitter sebagai tanda keanggotaan dan hak akses ke halaman pencarian Twitter. Hasil pengambilan data dari Twitter kemudian disimpan dalam format *file* Excel. Gambar 3 menunjukkan contoh data yang diperoleh dari proses pengambilan data tersebut.

created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id_str	conversation_id_str	username	tweet_url
Thu Dec 3 13:4E+18	@diptyas		0	1	0	0	in	8.62E+17	1.34E+18	lady_mi	https://twitter.com/lady_mi/status/134E+18
Thu Dec 3 13:4E+18	Pandemi		0	0	0	0	in	94275536	1.34E+18	rdtlco	https://twitter.com/rdtlco/status/134E+18
Thu Dec 3 13:4E+18	2. Gajiny		0	1	0	4	in	1.02E+18	1.34E+18	salima252	https://twitter.com/salima252/status/134E+18
Thu Dec 3 13:4E+18	SELMA	Ler	0	0	0	5	in	7.60E+17	1.34E+18	kumparan	https://twitter.com/kumparan/status/134E+18
Thu Dec 3 13:4E+18	Menolak		0	0	0	0	in	1.22E+18	1.34E+18	bregaswai	https://twitter.com/bregaswai/status/134E+18
Thu Dec 3 13:4E+18	@tubirfes		0	0	0	0	in	1.22E+18	1.34E+18	boengame	https://twitter.com/boengame/status/134E+18
Thu Dec 3 13:4E+18	Alasan	Kai	0	0	0	0	in	1.29E+18	1.34E+18	niko_fend	https://twitter.com/niko_fend/status/134E+18
Thu Dec 3 13:4E+18	@bayurizl		0	1	0	0	in	5.07E+08	1.34E+18	sibabang	https://twitter.com/sibabang/status/134E+18
Thu Dec 3 13:4E+18	@tubirfes		0	1	0	0	in	2.22E+09	1.34E+18	Tatafurma	https://twitter.com/Tatafurma/status/134E+18
Thu Dec 3 13:4E+18	@Rachma		0	0	0	0	in	3.54E+08	1.34E+18	Hellompu	https://twitter.com/Hellompu/status/134E+18
Thu Dec 3 13:4E+18	@ojkindo		0	0	0	0	in	1.27E+18	1.34E+18	RAAdilestai	https://twitter.com/RAAdilestai/status/134E+18
Thu Dec 3 13:4E+18	@DAMEC		0	0	0	0	in	1.28E+18	1.34E+18	heeseung	https://twitter.com/heeseung/status/134E+18
Thu Dec 3 13:4E+18	@tubirfes		0	0	0	0	in	8.91E+17	1.34E+18	mvikvkkdi	https://twitter.com/mvikvkkdi/status/134E+18

Gambar 3. Contoh Data Hasil Scapping

Terdapat beberapa kolom yaitu:

- created_at, adalah waktu (hari, tanggal, bulan, tahun dan jam) tweet dibuat oleh user
- id_str, nomor identitas dari tweet (Post Identifier (*integer casted inside string*))
- full_text, adalah isi tweet yang dibuat oleh user contoh: “@renayku Bener banget, kak! Asuransi itu sangat penting, karena kalau beban biaya kesehatan ditanggung sendiri itu berat. Dulu aku nggak percaya sama the power of asuransi, tapi setelah ada temen yg share pengalamannya saat sakit, lsg buru2 bikin asuransi. Aku pake AXA pun loh.”
- quote_count, banyaknya tweet dikutip (*quoted*) oleh user lain
- reply_count, banyaknya tweet dibalas oleh user lain
- retweet_count, banyaknya tweet di *re-tweet* oleh user lain
- favorite_count, banyaknya tweet disukai user lain
- lang, bahasa yang digunakan dalam tweet. Penelitian ini memilih tweet yang berbahasa Indonesia (in)
- user_id_str, nomor identitas dari user Twitter
- conversation_id_str, nomor identitas percakapan (jika ada *reply*)
- username, nama user dari pemilik tweet
- tweet_url, alamat URL dari tweet

Tidak semua data tersebut dipakai pada tahapan selanjutnya. Hanya kolom full_text yang akan digunakan. Karena penelitian ini berfokus dalam menentukan sentimen pada tweet. Dari hasil *scapping* didapat data sebanyak 35.190 secara lengkap terdapat pada Tabel 2.

Tabel 2. Total Data Scapping

	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Agt	Sept	Okt	Nov	Des	TOTAL
2020	672	769	926	728	1157	981	1078	1042	678	691	545	608	9875
2021	868	592	562	795	477	603	711	621	574	1122	943	86	7954
2022	624	1147	749	468	501	1092	551	482	713	1186	955	708	9176
2023	1065	990	867	846	1086	792	849	875	815				8185
Grand Total													35190

Dari 35.190 data hasil *scapping*, tidak semua berupa pendapat pada asuransi kesehatan, banyak yang berupa promosi asuransi kesehatan, edukasi, *podcast*, berita, penggalangan dana, pengalaman user saat di luar negeri, lowongan kerja, fasilitas yang didapatkan dari kantor dan konten tentang rencana keuangan. tweet yang berisi hal tersebut tidak dimasukkan dalam analisis. Jumlah hasil data seleksi yang berisi *review*/testimoni terdapat pada Tabel 3.

Tabel 3. Total Data Scapping

Tahun	Total data hasil scapping	Twit yang berisi review/testimoni
2020	9875	314
2021	7954	512
2022	9176	304
2023	8185	353
Grand Total	35190	1483

3.2 Praproses Data

Sebanyak 1.483 data tweet yang telah dipilih akan dipraproses terlebih dahulu sebelum digunakan untuk algoritme Naïve Bayes. Berikut adalah tahapan untuk praproses data tweet:

1. *Casefolding* (mengubah tweet ke dalam huruf kecil), menghapus karakter yang bukan huruf (termasuk angka juga dihapus), URL, *emoticon*, *hashtag* dan *mention*. Data sebelum melewati praproses terdapat pada Gambar 4.

index	full_text
0	@sabdaps Bpjs adalah asuransi kesehatan dengan coverage yang mantab betul! Not perfect, but almost.
1	@wikusabira Yeps, betul banget ini BPJS Tenaga Kerja juga bantu banget saat saya di PHK waktu pandemic, BPJS Kesehatan banyak bantu oma saya saat berobat katarak sesaat sebelum dia meninggal.
2	Hal yang membangongkan adalah aku bayar asuransi kesehatan swasta yang mahal tapi tidak bisa dipake berobat jalan. Alias mahal benar obatnya kalau mandiri begini, bpjs antrinya bukan main
3	@selphieusagi Iya, asuransi kesehatan emang penting sih. Minimal bpjs. kalau bisa yang lain syukur
4	@eskriiiiiimm @selphieusagi Tergantung paket asuransi kesehatan yang diambil kak. kalau aku pake Pruden full cover untuk RS Indonesia 550 ribu/bln. Ambilnya waktu usia 22 tahun (cover sampai 55 tahun) dan mmg tidak ada riwayat sakit. jadi biaya premi tergantung banyak faktor. tahun lalu covid masuk 10 hari 45 juta dicover full

Gambar 4. Data Sebelum Praproses

Adapun data yang telah melewati tahap praproses terdapat pada Gambar 5.

index	full_text
0	bpjs adalah asuransi kesehatan dengan coverage yang mantab betul not perfect but almost
1	yeps betul banget ini bpjs tenaga kerja juga bantu banget saat saya di phk waktu pandemic bpjs kesehatan banyak bantu oma saya saat berobat katarak sesaat sebelum dia meninggal
2	hal yang membangongkan adalah aku bayar asuransi kesehatan swasta yang mahal tapi tidak bisa dipake berobat jalan alias mahal benar obatnya kalau mandiri begini bpjs antrinya bukan main
3	iya asuransi kesehatan emang penting sih minimal bpjs kalau bisa yang lain syukur
4	tergantung paket asuransi kesehatan yang diambil kak kalau aku pake pruden full cover untuk rs indonesia ribu/bln ambilnya waktu usia tahun cover sampai tahun dan mmg tidak ada riwayat sakit jadi biaya premi tergantung banyak faktor tahun lalu covid masuk hari juta dicover full
5	kemaren masuk rs tagihan juta cuma bayar rb pentingnya asuransi kesehatan
6	bayangkan operasi kista rpjuta biayanya untung di tanggung asuransi kesehatan murni miultimate healthcare manulife langsung pake kartu bebas pilih rumah sakit yang alatnya bagus bebas pilih dokter yang sudah berpengalamanjam terbang tinggi handle pasien

Gambar 5. Data Setelah Praproses

2. *Tokenizing*, yaitu memisahkan kalimat menjadi kata. Setiap dideteksi spasi, maka kata dipisahkan. Hasil *tokenizing* terdapat pada Gambar 6.

index	full_text,Label
0	" ,bpjs,adalah,asuransi,kesehatan,dengan,coverage,yang,mantab,betul,not,perfect,but,almost"
1	" ,yeps,betul,banget,ini,,bpjs,tenaga,kerja,juga,bantu,banget,saat,saya,di,phk,waktu,pandemic,bpjs,kesehatan,ban yak,bantu,oma,saya,saat,berobat,katarak,, ,sesaat,sebelum,dia,meninggal"
2	"hal,yang,membangongkan,adalah,aku,bayar,asuransi,kesehatan,swasta,yang,mahal,tapi,tidak,bisa,dipake,berobat jalan,alias,mahal,benar,obatnya,kalau,mandiri,begini,bpjs,antrinya,bukan,main"
3	" ,iya,asuransi,kesehatan,emang,penting,sih,minimal,bpjs,kalau,bisa,yang,lain,syukur"
4	" , ,tergantung,paket,asuransi,kesehatan,yang,diambil,kak,kalau,aku,pake,pruden,full,cover,untuk,rs,indonesia,,rib u/bln,ambilnya,waktu,usia,,tahun,cover,sampai,,tahun,dan,mmg,tidak,ada,riwayat,sakit,jadi,biaya,premi,tergantu ng,banyak,faktor,tahun,lalu,covid,masuk,,hari,,juta,dicover,full"
5	" ,kemaren,masuk,rs,tagihan,juta,cuma,bayar,rb,pentingnya,asuransi,kesehatan"

Gambar 6. Data Tokenizing

3. Langkah selanjutnya adalah membuang *stopword*. *Stopword* adalah kata yang tidak memiliki makna. Contoh *stopword* dalam Bahasa Indonesia adalah yang, ke, di, dari, adalah, dsb. Setelah *stopword* dibuang, dilakukan proses *stemming* yaitu mengubah kata yang berimbuhan menjadi kata dasar. Pada penelitian ini, digunakan *library* Sastrawi dari Python. Sastrawi Python adalah *library* dari Python yang memungkinkan kita mengubah kata berimbuhan menjadi kata dasarnya (dalam bahasa Indonesia). Hasil langkah ini, terdapat pada Gambar 7.

	full_text	Label
0	" ,bpjs,asuransi,kesehatan,coverage,mantab,not,perfect,but,almost"	Positif
1	" ,yeps,banget,,bpjs,tenaga,kerja,bantu,banget,phk,pandemic,bpjs,kesehatan,bantu,oma,berobat,"	Positif
2	"membangungkan,bayar,asuransi,kesehatan,swasta,mahal,dipake,berobat,jalan,alias,mahal,obatn	Negatif
3	" ,iya,asuransi,kesehatan,emang,sih,minimal,bpjs,syukur"	Positif
4	" ,,tergantung,paket,asuransi,kesehatan,diambil,kak,pake,pruden,full,cover,rs,indonesia,,ribu/bln,	Positif
5	" ,kemaren,masuk,rs,tagihan,juta,bayar,rb,asuransi,kesehatan"	Positif

Gambar 7. Data setelah dibuang *stopword*-nya dan dilakukan *stemming*

4. Data hasil *stemming* kemudian dihitung bobotnya menggunakan *tf-idf* (*term frequency–inverse document frequency*). *Tf-idf* adalah ukuran statistik yang menggambarkan pentingnya suatu istilah terhadap sebuah dokumen. Semakin sering sebuah kata muncul dalam dokumen, maka semakin penting kata tersebut. Kata yang memiliki bobot tinggi, bisa dijadikan “ciri” kelas tertentu. Misal jika kata “syukur” tinggi bobotnya tinggi di kelas positif, maka kata “syukur” merupakan kata khas atau ciri dari kelas positif. Hasil *tf-idf* untuk tweet nomor 0 dan 1 terdapat pada Gambar 8.

(0, 117)	0.4150432686422715	(1, 3215)	0.25699633507700287
(0, 823)	0.37512332963171907	(1, 2285)	0.2912981132260146
(0, 4039)	0.43839493599452023	(1, 566)	0.18619396604405788
(0, 3636)	0.3984749969839678	(1, 3713)	0.3311612725859855
(0, 2947)	0.4150432686422715	(1, 3798)	0.3311612725859855
(0, 954)	0.38562364707028096	(1, 4103)	0.3010059566394192
(0, 2480)	0.05896314170187659	(1, 384)	0.43269856398124834
(0, 267)	0.05872581046749067	(1, 2463)	0.19975791513313387
(0, 724)	0.09031746605321972	(1, 5125)	0.2912981132260146
		(1, 368)	0.2494667145813633
		(1, 5571)	0.3311612725859855
		(1, 2480)	0.04454045299899475
		(1, 724)	0.13645069565903764

Gambar 8. Data hasil perhitungan bobot *tf-idf*

- (0,117) mempunyai arti angka 0 adalah indeks tweet . Angka 0 berarti tweet ke-0. Angka 117 adalah indeks kata (ditentukan *random* oleh mesin). Kemudian untuk yang (1,3215) itu berarti tweet ke-1 dan indeks kata ke-3215. Untuk angka setelah indeks, misal 0.41504, 0.375123 dan seterusnya, adalah bobot dari kata berdasarkan perhitungan *tf-idf*. Semakin tinggi nilai bobot, maka semakin penting kata tersebut.
5. Data kemudian dibagi menjadi data latih dan data uji. Sebanyak 80% data digunakan sebagai data latih untuk membuat model klasifikasi menggunakan *Naïve Bayes*. Adapun 20% data sisanya digunakan untuk menguji model klasifikasi sehingga diketahui akurasi. Data uji akan dimasukkan ke model klasifikasi kemudian dibandingkan kelas hasil prediksi (menggunakan algoritme klasifikasi) dengan kelas asli data uji. Jika sama, maka algoritme klasifikasi dapat memprediksi kelas dengan tepat. Jika berbeda maka kebalikannya, algoritme klasifikasi salah memprediksi kelas data uji.

3.3 Labeling Data

Data yang sudah diproses kemudian diolah lebih lanjut menggunakan pemrograman Python untuk diberi label sebagai sentimen positif atau negatif. Algoritme Python akan mengidentifikasi kata-kata yang merupakan penanda sentimen positif dan negatif. Misalnya, jika terdapat kata "berguna," itu akan dianggap sebagai sentimen positif, sementara jika ada kata "mahal," itu dianggap sebagai sentimen negatif. Namun, tidak semua hasil pelabelan dengan menggunakan pemrograman Python tersebut selalu tepat, karena terdapat beberapa tweet yang mengandung kata-kata positif dan negatif dalam satu kalimat, seperti contoh berikut:

“Walaupun aku kadang kesel sama sistem BPJS, tapi serius BPJS tuh berguna bgt, ga cuma BPJS sih, ya intinya asuransi kesehatan, yg namanya sakit orang kan gatau ya, yah kadang beberapa operasi/obat ga bisa ke cover, but at least kan sudah berusaha”

Oleh karena itu, pada penelitian ini dilakukan pengecekan kembali secara manual. Contoh hasil Labeling terdapat pada Tabel 4.

Beberapa contoh kata yang termasuk kelas positif dan negative terdapat pada Tabel 5.

Tabel 5. Beberapa Contoh Kata yang Termasuk Kelas Positif dan Negatif

Kata Positif	Kata Negatif
Asuransi	Asuransi swasta
BPJS Kesehatan	BPJS Kesehatan
Asuransi swasta	Bayar
Discover	Asuransi kesehatan
Asuransi	Wanda hamida
Keluarga	Beda
Fasilitas	Premi
Untung	Susah
Axa Mandiri	Usaha
Manfaat	Obat
Gratis	Perintah
Wajib	Korban

3.4 Klasifikasi dengan Naïve Bayes

Tahapan terakhir adalah klasifikasi menggunakan algoritme Naïve Bayes dan perhitungan akurasi. Klasifikasi *Naïve Bayes* dilakukan dengan menggunakan fungsi `MultinomialNB()`, adapun perhitungan akurasi menggunakan fungsi `confusion_matrix()`. Hasil akurasi pemrograman terdapat pada Gambar 11.

```
MultinomialNB Accuracy: 0.7441077441077442
MultinomialNB Precision: 0.8913043478260869
MultinomialNB Recall: 0.36607142857142855
MultinomialNB f1_score: 0.5189873417721518
confusion matrix:
[[ 41  71]
 [  5 180]]
-----
              precision    recall  f1-score   support

   Negatif      0.89      0.37      0.52      112
   Positif      0.72      0.97      0.83      185

 accuracy              0.74      297
 macro avg              0.80      297
 weighted avg           0.78      297
```

Gambar 11 Evaluasi Model

Dari tahapan-tahapan tersebut, didapatkan akurasi dari algoritme Naïve Bayes adalah 74.41%. Tingkat akurasi algoritme Naïve Bayes belum cukup baik untuk menduga sentimen *user* Twitter terhadap penyedia layanan asuransi kesehatan. Diduga karena adanya kalimat positif dan negatif yang ada dalam satu tweet (seperti yang telah dibahas pada tahap labeling), sehingga algoritme sulit mengidentifikasi *purely* positif atau negatif. Pada *word cloud* yang didapat juga Gambar 4 dan Gambar 5 diketahui ada beberapa kata yang masuk ke kedua kelas. Contoh: BPJS Kesehatan, Asuransi Kesehatan, Asuransi Swasta itu sering muncul di kelas positif dan negatif. Hal tersebut disebabkan karena banyak pro-kontra di dalam tweet, ada yang mendukung ada pula yang mencela. Contoh: “aku kesel sama sistem BPJS” di sisi lain, yang menggunakan review positif berkata:”tapi BPJS berguna banget”. Penggunaan singkatan, *typographical error* (salah ketik) dan penggunaan kata tidak baku juga bisa jadi merupakan penyebab kurangnya akurasi algoritme Naïve Bayes.

3.5 Pembahasan Analisis Data

Dari hasil analisis juga diketahui, hal-hal yang banyak dibicarakan terkait asuransi kesehatan pada tahun 2020-2023 terdapat pada Tabel 6.

Tabel 6. Hal-hal yang Dibicarakan

Tahun	Topik yang banyak dibicarakan
2020	<ol style="list-style-type: none"> 1. Berita korupsi BPJS 2. Vaksin Covid tidak bisa pake BPJS 3. Premi BPJS naik saat pandemi 4. Karyawan BPJS menggunakan asuransi swasta 5. Prokontra BPJS bangkrut (ada yang mengharapkan BPJS bangkrut, tapi ada juga yang mendukung adanya BPJS)
2021	<ol style="list-style-type: none"> 1. Wanda hamidah ditipu prudential 2. Anies baswedan menggrebek penyedia layanan asuransi equity terkait kebijakan WFH 3. Data BPJS bocor 4. Pro-kontra Kitabisa, OVO, Gojek meluncurkan asuransi kesehatan dengan premi 10-ribuan
2022	<ol style="list-style-type: none"> 1. pemberlakuan kelas rawat inap standar (KRJS) 2. pro-kontra kebijakan pemerintah mewajibkan BPJS (banyak yang keberatan harus membayar iuran BPJS setiap bulan padahal sudah punya asuransi swasta. Mereka inginnya BPJS tidak diwajibkan jika sudah memiliki asuransi kesehatan lain. Tapi banyak juga yang mendukung sistem “gotong-royong” di BPJS)
2023	<ol style="list-style-type: none"> 1. Bapak Prabowo banyak melakukan kampanye tentang asuransi kesehatan dalam rangka pilpres 2. BPJS belum melayani infertilitas 3. Meningkatnya klaim asuransi kesehatan karena ISPA 4. Berita mengenai istri Indra Bekti menggalang dana saat suaminya sakit (karena tidak mempunyai asuransi kesehatan) 5. karyawan BPJS menggunakan asuransi swasta 6. pro kontra BPJS halal-haram

Dari ringkasan hasil twit menurut pandangan masyarakat terkait BPJS Kesehatan dapat dilihat sisi positif dan negatifnya. Adapun *review* positif dari Masyarakat pengguna Twitter adalah:

1. BPJS Kesehatan dipilih karena mempunyai premi yang murah
2. BPJS Kesehatan mencakup berbagai jenis penyakit termasuk penyakit kritis dan operasi
3. Pendaftaran BPJS Kesehatan dapat dilakukan tanpa *medical check-up* terlebih dahulu, tidak mengenal *pre-existing condition* dan terdapat plafon yang tidak terbatas.

Adapun *review* negatif dari Masyarakat pengguna Twitter adalah:

1. Proses pengobatan dengan BPJS Kesehatan melibatkan prosedur yang rumit dan memerlukan waktu yang cukup lama untuk menunggu antrian.
2. BPJS Kesehatan membuat sulitnya proses penghentian kepesertaan karyawan swasta karena setiap pekerja diwajibkan memiliki BPJS. Banyak orang merasa kesulitan saat harus membayar premi ganda, untuk asuransi swasta maupun BPJS.
3. Nampaknya ada tuntutan agar setiap individu memiliki BPJS Kesehatan, seolah-olah menjadi prasyarat untuk transaksi jual-beli properti atau untuk pelaksanaan perjalanan Haji/Umroh.
4. BPJS Kesehatan seharusnya menjadi program gotong royong antara kelompok yang lebih mampu dan kelompok yang kurang mampu, namun ironisnya, justru masyarakat yang lebih mampu yang menanggung beban BPJS Kesehatan.

Dari ringkasan hasil twit menurut pandangan masyarakat terkait Asuransi Kesehatan Swasta dapat dilihat sisi positif dan negatifnya juga. Adapun *review* positif dari masyarakat pengguna Twitter adalah:

1. Asuransi swasta memiliki prosedur yang mudah dan antri lebih cepat
2. Terdapat banyak pilihan produk yang tersedia, dapat dipilih dan disesuaikan dengan penghasilan individu

Adapun negatif dari Masyarakat pengguna Twitter adalah:

1. Terdapat banyak tawaran melalui telepon, bahkan memberi kesan memaksa membeli produk karena dihubungi berulang kali
2. Banyaknya kasus gagal bayar penyedia layanan asuransi seperti asuransi Jiwasraya dan Bumi Putera
3. Adanya konsumen yang merasa ditipu, awalnya ditawari asuransi pendidikan namun beralih secara sepihak menjadi asuransi kesehatan
4. Masih ditagih pembayaran melalui *autodebet* rekening padahal masa polis sudah selesai
5. Berubahnya produk yang dibeli dari asuransi kesehatan menjadi asuransi jiwa tanpa pemberitahuan.

4. KESIMPULAN

Telah dilakukan *sentiment analysis* pada data Twitter dari *user* Indonesia yang mengandung kata kunci terkait asuransi kesehatan. Tujuan penelitian ini adalah untuk mengetahui pendapat masyarakat Indonesia terhadap penyedia layanan asuransi kesehatan melalui data *twit*. Media sosial Twitter dipilih, karena Twitter merupakan salah satu media sosial yang paling banyak digunakan di Indonesia. Pengguna Twitter di Indonesia mencapai 18,45 juta pada tahun 2022.

Sebanyak 35.190 data *twit* dikumpulkan mulai dari Januari 2020 sampai September 2023. Namun, tidak semua data *twit* tersebut relevan dengan tujuan penelitian, sehingga perlu dilakukan seleksi. Setelah dilakukan seleksi, didapat 1.483 data *twit* yang melalui tahapan praproses yaitu *tokenizing*, *filtering*, *stemming* dan *labeling*. Data yang sudah dipraproses kemudian digunakan untuk *sentiment analysis* menggunakan algoritme Naïve Bayes.

Didapatkan akurasi dari algoritme Naïve Bayes adalah 74.41%. Tingkat akurasi algoritme Naïve Bayes belum cukup baik untuk menduga sentimen *user* Twitter terhadap penyedia layanan asuransi kesehatan. Hal ini disebabkan karena adanya kalimat positif dan negatif yang ada dalam satu *twit*, sehingga algoritme sulit mengidentifikasi *purely* positif atau negatif. Pada *word cloud* yang didapat juga diketahui ada beberapa kata yang masuk kedalam kedua kelas (positif dan negatif). Hal tersebut disebabkan karena banyak pro-kontra di dalam *twit*, ada yang mendukung ada pula yang mencela. Penggunaan singkatan, *typographical error* (salah ketik) dan penggunaan kata tidak baku juga bisa menjadi penyebab kurangnya akurasi algoritme Naïve Bayes.

Diperlukan analisis lebih lanjut untuk meningkatkan akurasi *sentiment analysis* untuk mengetahui pendapat masyarakat pada penyedia layanan asuransi kesehatan ini. Diantaranya dengan menambah data uji sehingga data input semakin beragam. Melakukan tahap praproses lebih teliti seperti mengatasi singkatan, salah ketik dan kata tidak formal. Selain itu, karena Naïve Bayes hanya mengidentifikasi kata per kata, diperlukan algoritme lain seperti *sequential pattern mining* untuk mengidentifikasi pola kemunculan kata. Misal jika kata BPJS diikuti dengan kata kesal maka Negatif. Jika BPJS diikuti kata berguna maka Positif.

Dari hasil analisis juga diketahui hal-hal yang menyebabkan masyarakat berpendapat negatif pada layanan asuransi kesehatan, diantaranya: prosedur yang rumit dan memerlukan waktu yang cukup lama untuk menunggu antrian, pro-kontra wajib BPJS Kesehatan bagi karyawan swasta, *telemarketing* yang mengganggu, kasus gagal bayar dan lain sebagainya. Harapannya dari hasil analisis ini dapat dijadikan bahan evaluasi dan perbaikan layanan asuransi kesehatan.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini. Seluruh pimpinan STMA Trisakti yang telah mendukung dan menyediakan fasilitas penelitian. Tak luput pula kepada kepala P3M (Pusat Penelitian dan Pengabdian kepada Masyarakat) STMA Trisakti atas motivasi serta komitmen dalam memayungi penelitian. Tak lupa pula ucapan terima kasih kepada rekan-rekan sivitas akademik STMA Trisakti

REFERENCES

- [1] P. J. Ng, J. Rumengan, F. Fadlan, and I. Idham, "Eksistensi Otoritas Jasa Keuangan Dalam Memberikan Perlindungan Hukum Kepada Pemegang Polis Asuransi," *J. Ius Const.*, vol. 5, no. 2, p. 196, 2020, doi: 10.26623/jic.v5i2.2308.
- [2] N. Suyatna, "Effectiveness of OJK ' s Stimulus Policy Accounting of Qardh Funds on Solvency Level (During Covid-19 Pandemic) in The Sharia Life Insurance Sector in Indonesia," *J. Ekon.*, vol. 11, no. 01, pp. 689–696, 2022.
- [3] Badan Pusat Statistik, "Jumlah Penduduk Yang Dicakup Asuransi Kesehatan Atau Sistem Kesehatan Masyarakat per 1000 Penduduk, 2014 - 2018," 2019. https://bps.go.id/dynamicstable/2018/07/02_12:58:57.544653/1513/jumlah-penduduk-yang-dicakup-asuransi-kesehatan-atau-sistem-kesehatan-masyarakat-per-1000-penduduk.html#:~:text=Jumlah Penduduk Yang Dicakup Asuransi Kesehatan Atau Sistem,ini. Badan Pusat Sta.
- [4] H. Jasin, S. Mujiatun, M. Fauzi Rambe, and R. Bahagia Siregar, "Apakah Kepercayaan Memediasi Pengaruh Reputasi Bank dan Religiusitas Terhadap Purchase Intention?," *J. Ilm. Manaj. dan Bisnis*, vol. 22, no. 1, pp. 86–102, 2021,

- doi: 10.30596/jimb.v22i1.5630.
- [5] M. Ahmadi, H. Hairul, and K. Kurniaty, "Analisis Inovasi Fitur Pelayanan Pada Aplikasi Dalam Meningkatkan Kualitas Kinerja Driver Ojek Online PT. GOJEK INDONESIA BANJARMASIN." Universitas Islam Kalimantan, 2019.
- [6] C. Destitus, W. Wella, and S. Suryasari, "Support Vector Machine VS Information Gain: Analisis Sentimen Cyberbullying di Twitter Indonesia," *Ultim. InfoSys J. Ilmu Sist. Inf.*, vol. 11, no. 2, pp. 107–111, 2020.
- [7] H. Priyatna and M. L. Fatahillah, *Kamus Teknologi Informasi dan Komunikasi*. Nuansa Cendekia, 2023.
- [8] M. Tanninen, "Contested technology: Social scientific perspectives of behaviour-based insurance," *Big Data Soc.*, vol. 7, no. 2, p. 2053951720942536, 2020.
- [9] M. Qazi, K. Tollas, T. Kanchinadam, J. Bockhorst, and G. Fung, "Designing and deploying insurance recommender systems using machine learning," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 10, no. 4, pp. 1–33, 2020, doi: 10.1002/widm.1363.
- [10] D. Kurnianingrum, N. Nugraha, D. Disman, B. S. Purnomo, and M. Karmagatri, "The Role of Twitter in Business, Economics, and Finance Research: a Bibliometric Analysis," *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 15, pp. 6092–6113, 2023.
- [11] I. A. Asqolani and E. B. Setiawan, "Hybrid Deep Learning Approach and Word2Vec Feature Expansion for Cyberbullying Detection on Indonesian Twitter.," *Ingénierie des Systèmes d'Information*, vol. 28, no. 4, 2023.
- [12] M. Ahlgren, "STATISTIK TWITTER, FAKTA & TREN UNTUK 2023," *Websiterating*. <https://www.websiterating.com/id/research/twitter-statistics/>.
- [13] S. Liu and S. D. Young, "A survey of social media data analysis for physical activity surveillance," *J. Forensic Leg. Med.*, vol. 57, pp. 33–36, 2018.
- [14] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.
- [15] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes," *Inf.*, vol. 12, no. 5, 2021, doi: 10.3390/info12050204.
- [16] W. Wagiran, "Metodologi penelitian pendidikan: Teori dan implementasi," *Yogyakarta Budi Utama*, 2013.
- [17] L. P. Mahyuni, *Strategi Praktis Penelitian dan Penulisan Karya Ilmiah Untuk Sukses Publikasi Pada Jurnal Bereputasi*. Syiah Kuala University Press, 2021.
- [18] E. Yuniar, D. S. Utsalinah, and D. Wahyuningsih, "Implementasi Scrapping Data Untuk Sentiment Analysis Pengguna Dompot Digital dengan Menggunakan Algoritma Machine Learning," *J. Janitra Inform. dan Sist. Inf.*, vol. 2, no. 1, pp. 35–42, 2022, doi: 10.25008/janitra.v2i1.145.