

ANALISIS SENTIMEN MASYARAKAT TERHADAP PROGRAM BADAN PENYELENGGARA JAMINAN SOSIAL (BPJS) MENGUNAKAN DATA TWITTER

Zefanya Ardika*¹⁾, Alz Danny Wowor²⁾

1. Universitas Kristen Satya Wacana, Salatiga, Indonesia
2. Universitas Kristen Satya Wacana, Salatiga, Indonesia

Article Info

Kata Kunci: BPJS, *Classification*; *Confusion Matrix*; *Data Mining*; *Sentiment*.

Keywords: BPJS; *Classification*; *Confusion Matrix*; *Data Mining*; *Sentiment*.

Article history:

Received 9 November 2023

Revised 23 November 2023

Accepted 7 December 2023

Available online 1 March 2024

DOI :

<https://doi.org/10.29100/jipi.v9i1.4272>

* Corresponding author.

Zefanya Ardika

E-mail address:

672019225@student.uksw.edu

ABSTRAK

Jaminan Kesehatan Nasional merupakan salah satu program kesehatan pemerintah. Program ini menggunakan sistem premi, mirip dengan asuransi kesehatan pada umumnya, dimana untuk memastikan bahwa setiap orang dapat mengakses pelayanan kesehatan yang merata dan adil. Program ini adalah salah satu cara kebijakan ini dipraktikkan dengan mendirikan Badan. Dengan Badan hukum yakni Badan Penyelenggara Jaminan Sosial (BPJS) yang terbentuk untuk menyelenggarakan program jaminan kesehatan. Riset ini dibuat untuk melihat sentimen yang diberikan oleh masyarakat terkait program BPJS, riset dilakukan dengan Analisa Sentimen menggunakan metode atau model *Naive Bayes*, SVM, dan *Random Forest*. Data yang digunakan dalam riset ini merupakan Data tweet dari pengguna *twitter* yang diambil pada April 2023 dengan menggunakan API Twitter, mengambil 6000 unggahan *tweet* pengguna. Riset ini menggunakan model *Naive Bayes*, SVM, dan *Random Forest* yang masing masing memberikan hasil yang cukup akurat dengan tingkat akurasi sebesar 99.6%, 97.6%, dan 98.5%. Dari 1053 data bersih setelah *preprocessing*, menghasilkan metode terbaik diantara metode lain dengan metode SVM menghasilkan 339 *tweet* positif. 290 *tweet* netral, dan 211 *tweet* negatif.

ABSTRACT

The National Health Insurance is one of the government's health programs. This program uses a premium system, like health insurance in general, which ensures that everyone can access health services that are equitable and fair. This program is one of the ways this policy is put into practice by establishing the Agency. With a legal entity, namely the Badan Penyelenggara Jaminan Sosial (BPJS) which was formed to administer the health insurance program. This research was made to see the sentiments given by the public regarding the BPJS program. The research was carried out by Sentiment Analysis using the Naïve Bayes method or model, SVM, and Random Forest. The data used in this research is tweet data from Twitter users taken in April 2023 using the Twitter API, taking 6000 user tweets. This research uses the Naïve Bayes, SVM, and Random Forest models, each of which gives quite accurate results with an accuracy rate of 99.6%, 97.6%, and 98.5%. Of the 1053 clean data after preprocessing, the results using the SVM method produce 339 positive tweets. 290 neutral tweets, and 211 negative tweets

I. PENDAHULUAN

JAMINAN Kesehatan Nasional merupakan salah satu program kesehatan pemerintah. Program ini menggunakan sistem premi, mirip dengan asuransi kesehatan pada umumnya, dimana untuk memastikan bahwa setiap orang dapat mengakses pelayanan kesehatan yang merata dan adil. Program ini adalah salah satu cara kebijakan ini dipraktikkan dengan mendirikan Badan [1]. dengan Badan hukum yakni Badan Penyelenggara Jaminan Sosial (BPJS) yang terbentuk untuk menyelenggarakan program jaminan Kesehatan [2]. Berdasarkan Peraturan Presiden No. 12 tahun 2013 pasal 6 ayat (1), kesertaan jaminan kesehatan bersifat wajib bagi seluruh penduduk Indonesia. Walaupun peraturan pemerintah tersebut telah ditetapkan masih saja di temukan keluarga yang belum menjadi peserta BPJS Kesehatan [3]. Pada dasarnya BPJS sungguh membantu karena salah satu manfaat dari BPJS itu

sendiri yakni memberikan layanan yang baik untuk peserta dalam lingkup kesehatan. Akan tetapi semakin banyak penduduk yang memakai layanan tersebut, maka semakin banyak juga pro dan kontra dari masyarakat Republik Indonesia[4]. Mulai dari yang tidak merasa terberatkan oleh iuran sampai yang merasa tidak diprioritaskan.

Permasalahan yang ada saat ini yaitu timbulnya permasalahan dimana di suatu wilayah ternyata pasien BPJS diterlantarkan oleh rumah sakit tertentu, lalu timbulnya masalah bahwa banyak masyarakat yang kurang mengerti mengenai prosedur dan standar pelayanan yang berhak didapatnya Ketika menggunakan fasilitas BPJS [5]. Dikutip melalui Kompas.id bahwa penggunaan layanan BPJS ini mengalami keterbatasan untuk klaim obat, dengan kata lain tidak semua obat bisa untuk didapatkan dan jika memang diluar pertanggung jawaban BPJS maka diharuskan untuk membeli dengan uang sendiri, serta terdapat pembagian kelas untuk beberapa kategorinya, seperti layanan rawat inap yang hanya bisa dinikmati oleh penerima BPJS kelas 1 saja, hal ini bertolak belakang dengan perusahaan asuransi lain yang dapat klaim seluruh layanan di rumah sakit termasuk akses untuk klaim obat-obatan [6]. Namun di samping itu, berdasarkan hasil wawancara kepada pasien BPJS dan Non BPJS bahwa pegawai sudah memberikan pelayanan yang cepat dengan pasien melengkapi persyaratan dan mengikuti alur pelayanan yang ada maka proses pelayanan akan cepat dan pelayanan disesuaikan dengan nomor antrean, persyaratan yang diberikan mudah dan tidak berbelit-belit [7].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan agar dapat mencari tahu bagaimana sentimen dari masyarakat khususnya pengguna media sosial *twitter* dalam memberi opini tentang keberlangsungan program BPJS yang ada selama ini, sehingga dapat menjadi suatu acuan untuk masyarakat dan BPJS untuk melihat apakah masih perlu peningkatan layanan yang ada atau bahkan sudah memuaskan. Dengan sentimen analisis dapat mendapatkan informasi yang dibutuhkan dari data yang mentah yang tidak terstruktur dan nantinya akan diolah, sehingga diharapkan dengan penelitian ini dapat diketahui bagaimana sentimen pengguna *twitter* terhadap program BPJS. Penelitian ini juga diharapkan dapat untuk mengetahui penggunaan *feature selection* terhadap akurasi proses analisis yang dilakukan. *Twitter* merupakan salah satu *platform* media sosial yang sangat terkenal di kalangan pengguna internet, hal ini disebabkan kesederhanaan dan kemudahan dalam pemakaiannya, serta pengguna dapat dengan bebas mengeluarkan opini mereka [8]. Dengan tujuan untuk mengetahui seberapa pro dan kontra program BPJS ini di kalangan masyarakat Indonesia khususnya pada media sosial *twitter* dan menjadikan evaluasi untuk BPJS dalam meningkatkan layanan kepada masyarakat.

Pada penelitian terdahulu yang diteliti oleh Anam Menjelaskan bahwa penggunaan algoritma SNA dan SVM untuk mendeteksi sentimen publik di Twitter, serta mendapatkan akurasi 92% menggunakan algoritma ini, telah tercapai. Namun, bukan berarti keduanya buruk karena SVM dengan pemilihan fitur GA dan PSO memiliki akurasi yang rendah dibandingkan dengan SVM. Hanya data split tunggal, yaitu 70:30 yang akan digunakan untuk pengujian [9].

Pada penelitian sebelumnya menyatakan bahwa metode *Support Vector Machine* (SVM) untuk Analisis Sentimen Masyarakat Program Kartu Prakerja di *Twitter*. Emosi yang digunakan adalah netral, negatif, dan positif. *Support Vector Machine* (SVM) adalah teknik untuk menganalisis opini publik dengan menggunakan data dari media sosial *Twitter*. Dengan menggunakan metode *Confusion Matrix*, nilai efektivitas klasifikasi SVM untuk sementara. Dalam penelitian ini, perbandingan dua *kernel-linier* dan RBF dilakukan. Berdasarkan hasil evaluasi nilai akurasi kernel linier menghasilkan nilai sebesar 98,67%, sedangkan kernel RBF menghasilkan nilai sebesar 98,34% [10]. Berdasarkan hasil penelitian tersebut menjadi pertimbangan penulis untuk melakukan penelitian dengan menggunakan metode klasifikasi SVM.

Pada penelitian yang diteliti oleh Samsir menjelaskan mengenai menganalisis opini publik terhadap pembelajaran daring pada masa pandemi COVID-19 di Indonesia pada awal November 2020. Penelitian dilakukan dengan penambahan teks berbasis dokumen pada *Twitter* yang dianalisis menggunakan algoritma *Naïve Bayes*. Temuan menunjukkan bahwa pembelajaran daring memiliki 30% sentimen positif, 69% sentimen negatif, dan 1% netral pada periode tersebut. Tingginya sentimen negatif dihasilkan karena ketidakpuasan masyarakat terhadap pembelajaran daring. Beberapa *tweet* menunjukkan kekecewaan dengan kata “stres” dan “malas” merupakan kata yang memiliki frekuensi tinggi dalam percakapan [11].

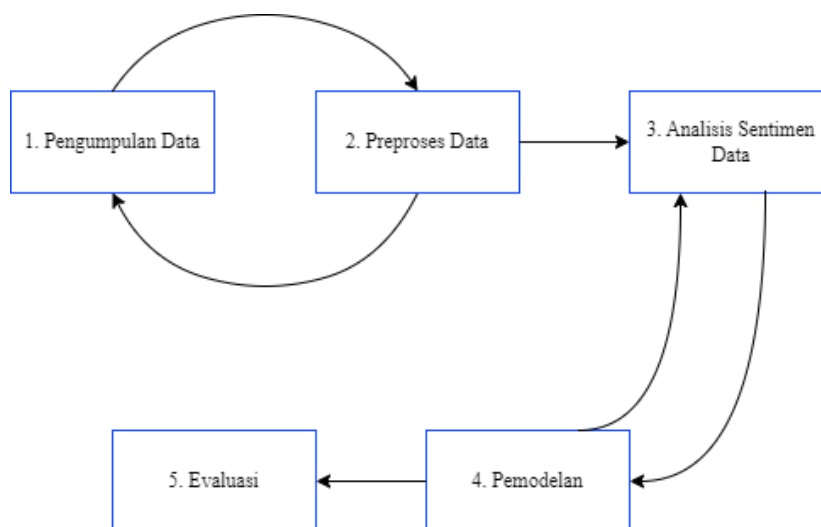
Berdasarkan penelitian sebelumnya, dinyatakan bahwa penggunaan metode *Random Forest* dalam mengklasifikasikan deteksi sarkasme pada proses analisis sentimen publik pada *platform Twitter*. Penelitian tersebut bertujuan untuk mendeteksi *tweet* yang bernilai positif akan dilakukan proses pendeteksian sarkasme menggunakan metode *Random Forest*. Data *training* yang dipakai untuk penelitian ini didasarkan pada penilaian nyata oleh pengguna *Twitter* terhadap tanggapan atau komentar pada suatu *tweet*, dengan dilakukan tahapan ekstraksi fitur dengan 4 set fitur yaitu *sentiment related*, *punctuation-relate*, *lexical and syntactic*, dan *pattern-*

relate sebagai tahapan ekstraksi. Hasil penelitian ini didapatkan nilai akurasi yang cukup tinggi sebesar 77,22 %, dengan jumlah data sebanyak 2.027 terhadap tweet pengguna. Hal ini disebabkan karena data training yang digunakan dan adanya beberapa data *noise* yang mempengaruhi performa metode Forest [18]. Oleh Sebab itu, penulis memilih untuk memasukkan Random Forest terkait Program Badan Penyelenggara Jaminan Umum (BPSJ) melalui media *Twitter*.

Dengan demikian penelitian ini ditujukan untuk mengetahui bagaimana opini masyarakat terhadap program BPJS berdasarkan dengan sentimen positif, negatif, dan netral di lingkup Indonesia memanfaatkan “cuitan” yang ada pada platform media sosial *Twitter*. Dengan melakukan penelitian ini, diharapkan untuk dapat mendapat pemahaman yang lebih baik terkait opini masyarakat terhadap Program BPJS, serta memberikan pandangan yang memiliki harga bagi pemerintah terkhusus Program BPJS dalam pengembangan layanan dan inovasi kepada masyarakat. Melalui penelitian ini, data yang diambil melalui fitur *Application Programming Interface* (API) pada *twitter*, dengan pemanfaatan aplikasi *Orange for Data Mining* sebagai penggunaan metode. Data akan digali dengan menggunakan kata kunci yang berhubungan dengan Program BPJS, dan dengan data yang ada akan melalui tahap *preprocess*, dan akan difokuskan pada pengenalan tipe konten dari berbagai informasi *tweet* pada *tools Orange for Data Mining* dengan menerapkan metode klasifikasi.

II. METODE PENELITIAN

Dalam penelitian ini melakukan pengambilan data opini masyarakat melalui *twitter*, dan untuk mendapatkan opini masyarakat tersebut melalui *tweet* dengan menggunakan API yang telah disediakan oleh *Twitter*, untuk memudahkan pengambilannya maka dilakukan pengambilan dataset tersebut melalui *tools Orange for data mining* yang dimana melakukan pengumpulan data menggunakan kata kunci, dan pada penelitian ini menggunakan beberapa kata kunci seperti BPJS, BPJS Kesehatan, dan BPJS Ketenagakerjaan, sehingga menghasilkan data berdasarkan Content, Author, Date, dan atribut unik lainnya.



Gambar. 1. Alur Desain Penelitian

Gambar 2 menunjukkan alur penelitian, yang menjelaskan tentang proses awal hingga akhir pada penelitian ini secara garis besar dibagi jadi 5 (lima) tahapan yakni:

A. Pengumpulan Data

Melalui API *Twitter* dengan menggunakan *beartoken* yang tersedia pada web *twitter for developer* menggunakan kata kunci BPJS, BPJS Ketenagakerjaan, BPJS Kesehatan dengan data yang diambil sebanyak 6000 data yang diambil dari *twitter* yang dimana menghasilkan data mentah dari opini masyarakat yang ada pada platform *twitter*.

B. Preproses Data

Text Preprocessing merupakan tahap awal pengolahan data yang mengubah teks menjadi data sehingga dapat diproses lebih lanjut. Algoritme klasifikasi tidak dapat menganalisis teks secara langsung, sebagai hasilnya, text preprocessing diperlukan untuk mempersiapkan teks untuk diubah menjadi data numerik. Teks harus dibagi, ada banyak cara untuk melakukan ini, bab, subbab, paragraf, frasa, dan akhirnya potongan kata atau token semuanya dapat dipisahkan dari sebuah dokumen. Selain itu, pada titik ini setiap angka, huruf kapital, atau karakter lain

diubah atau dihilangkan. Tahap-tahap Text Preprocessing yang pertama yaitu *Cleansing*, *cleansing* adalah menghilangkan tanda baca dan kata-kata yang tidak perlu disebut membersihkan. *Cleansing* dilakukan dengan tujuan mencegah gangguan selama proses klasifikasi. Kedua adalah *Case Folding* adalah proses untuk membuat proses selanjutnya lebih mudah, proses case folding adalah mengubah bentuk huruf menjadi seragam, apakah itu huruf kapital atau kecil. Yang ketiga yaitu *Tokenizing* atau tokenisasi adalah pemenggalan kalimat menjadi token atau potongan kata-kata disebut "tokenisasi". Ini memudahkan proses *stopword* dan *stemming*. Dan yang keempat adalah *Stemming* atau kata-kata yang sering ditemukan di dokumen sering memiliki banyak varian morfologik. Akibatnya, setiap kata yang bukan *stopword* direduksi ke akar kata yang sesuai; dengan kata lain, awalan atau akhiran dibuang dari kata untuk memperoleh bentuk akarnya. Ini menghasilkan kelompok kata yang cocok yang memiliki sintaksis yang berbeda dan hanya mengandung satu kata per kelompok.

C. Analisis Sentimen Data

Setelah melalui pemrosesan data sampai dengan analisa data, maka akan didapatkan hasil berupa numeric yang menyatakan bahwa suatu tweet memiliki sentimen positif, negatif, atau netral, dengan luaran numeric berupa angka tersebut maka dapat dilakukan labelling pada data yang ada untuk dapat dimodelkan menggunakan metode yang ada.

D. Pemodelan

Lalu melalui hasil tersebut dapat dimodelkan menggunakan pendekatan metode atau algoritma, dalam penelitian ini khususnya menggunakan beberapa metode, yaitu menggunakan SVM, *Naïve Bayes*, serta *Random Forest*. Dengan melalui pendekatan ketiga metode ini akan diketahui bahwa mana metode yang efisien atau tepat dengan precision sebagai acuan keakurasian penggunaan metode. Salah satu algoritma yang digunakan dalam metode klasifikasi adalah algoritma *Naïve Bayes*. Ini disederhanakan menjadi model probabilistik dan statistik dan berbasis pada teorema Bayes, yang menyatakan bahwa setiap atribut adalah bebas. Dengan kata lain, algoritma ini menganggap bahwa ciri kelas tertentu mungkin atau mungkin tidak memiliki hubungan dengan ciri kelas lainnya [14]. Metode ini sangat cocok digunakan sebagai pengklasifikasian sentimen pada Tugas Akhir ini dikarenakan memiliki beberapa kelebihan antara lain, sederhana, cepat, dan berakurasi tinggi. *Naïve Bayes* memiliki bentuk umum seperti berikut:

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

Gambar. 2. Perhitungan Rumus Naïve Bayes

Pada persamaan tersebut dapat dikatakan bahwa Peluang kejadian A sebagai B ditentukan dari peluang B saat A, dimana A merupakan hipotesis data B adalah kelas khusus, B merupakan data dengan kelas tidak dikenal, lalu $P(A|B)$ adalah probabilitas hipotesis A didasarkan pada kondisi B, $P(A)$ merupakan probabilitas hipotesis A, $P(B|A)$ merupakan probabilitas hipotesis B didasarkan pada kondisi A, dan $P(B)$ merupakan probabilitas B dari data yang ada.

Algoritma SVM, algoritma ini dapat mengubah data pelatihan awal ke dimensi yang lebih besar dengan menggunakan pemetaan nonlinier. Dalam hal ini, akan mencari *hyperplane* untuk memisahkan secara linier untuk dimensi baru, dan dengan pemetaan *nonlinier* yang tepat ke dimensi yang lebih besar, data dari dua kelas atau lebih selalu dapat dipisahkan dengan *hyperplane* tersebut [15]. Dimana dalam kasus ini garis tersebut berperan memisahkan *tweet* bersentimen positif (berlabel positif), *tweet* bersentimen netral (berlabel netral), dan *tweet* bersentimen negatif (berlabel negatif). Dalam penelitian ini, data yang dimasukkan dengan representasi vektor yang dihasilkan dari proses pembobotan. Setelah pelatihan klasifikasi SVM selesai, akan dihasilkan pola atau nilai yang akan digunakan dalam proses pengujian untuk proses pengujian SVM yang bertujuan untuk memberi label sentuhan pada *tweet*.

Random Forest (RF) merupakan salah satu metode machine learning yang dimanfaatkan untuk klasifikasi data yang memiliki jumlah banyak dan juga merupakan pengembangann metode *Classification and Regression Tree* (CART). Metode CART digunakan untuk membuat pohon keputusan, sehingga pohon keputusan tersebut tumbuh mencapai ukuran maksimum dan tidak dipangkas sehingga dihasilkan kumpulan pohon yang kemudian disebut forest. Hasil akhir klasifikasi metode *Random Forest* diambil dari *major voting* dari pohon yang terbentuk. Pada penelitian ini digunakan random forest classifier untuk memprediksi teks masuk ke kelas positif, negatif, atau netral [17]. Random Forest dalam data klasifikasi harus tahu bahwa akan sering menggunakan indeks Gini, atau rumus yang digunakan untuk memutuskan bagaimana simpul pada cabang pohon keputusan sebagai berikut:

$$Gini = 1 - \sum_{i=1}^c (P_i)^2$$

Gambar. 3. Pehitungan Rumus *Random Forest*

Rumus ini menggunakan kelas dan probabilitas untuk menentukan Gini dari setiap cabang pada sebuah node, menentukan cabang mana yang lebih mungkin terjadi. Di sini P_i mewakili frekuensi relatif kelas yang diamati dalam dataset dan c mewakili jumlah kelas.

E. Evaluasi

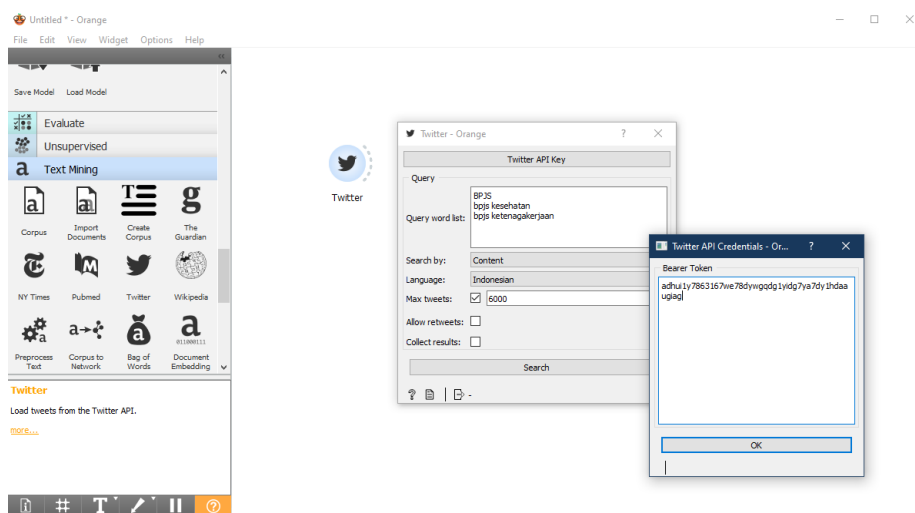
Mengklasifikasikan data yang sudah dianalisa sentimennya serta sudah didapatkan hasil keakurasian tiap metode, maka baru akan dapat diklasifikasi bahwa suatu tweet dari masyarakat merupakan sentimen positif, negatif, atau netral. Dengan visualisasi data yang ada dapat mempermudah dalam penentuan hasil akhir bagaimana sentimen masyarakat pada media Twitter.

III. HASIL PEMBAHASAN

Dalam penelitian ini, melakukan Analisa opini masyarakat khususnya pada *twitter* dengan pendekatan metode klasifikasi yaitu *Naïve Bayes*, *SVM*, dan *Random Forest* terhadap topik program Badan Penyelenggara Jaminan Sosial (BPJS). Oleh sebab itu maka pada bagian hasil dan pembahasan pada penelitian ini akan dijelaskan sebagai berikut:

A. Pengambilan Data

Dalam pengambilan data berupa opini masyarakat pada *twitter* ini membutuhkan penghubung agar data pada *twitter* dengan tools yang digunakan (*Orange Data mining*) dapat terhubung langsung, oleh karena itu dalam penelitian ini menggunakan *Bearer token* yang disediakan oleh *twitter*.



Gambar. 4. Pengambilan data *Twitter*

Pada gambar di atas menunjukkan cara untuk pengambilan data *twitter* berupa hasil *tweet* dari pengguna berdasarkan bahasa Indonesia dengan penggunaan beberapa kata kunci seperti BPJS, BPJS Ketenagakerjaan, BPJS Kesehatan, lalu dengan diambil data sebanyak 6000 tweet mentah yang dimana masih terdapat kata atau karakter yang tidak diperlukan, seperti emoji dan simbol simbol lainnya.

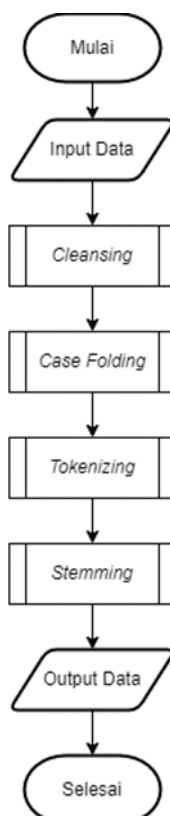
id	Content	Author	Date	Language	Location	Number of Likes	Number of Retweets	In Reply To	Author Name	Author Description	Author Tweets Count	Author Following Count	Author Followers Count	Author Listed Count	Author...
1	@soeyoto1 Inget za...	@SuhelanEla0	2023-05-06...	in	?	0	0	@soeyoto1	Suhelan ...	?	1741	34	5	0	Fal
2	RT @Restty_cayah...	@yadimanvasa	2023-05-06...	in	?	0	34	?	Djoeragan...	?	18375	159	173	0	Fal
3	Ganjar & Kade...	@GanjarKu	2023-05-06...	in	?	0	0	?	Ganjar Ku	Hai, jabatan jan...	1919	9	277	0	Fal
4	BPPS Last Kiss 800 ...	@karina_yasmine	2023-05-06...	ja	?	0	0	@karina_ya...	Yasmine K...	baru maen twit...	412	70	37	0	Fal
5	BPPS Last Kiss 800 ...	@karina_yasmine	2023-05-06...	ja	?	0	0	@karina_ya...	Yasmine K...	baru maen twit...	412	70	37	0	Fal
6	BPPS Last Kiss 800 ...	@karina_yasmine	2023-05-06...	ja	?	0	0	?	Yasmine K...	baru maen twit...	412	70	37	0	Fal
7	RT @DhifFeenBPP...	@dagiababoso	2023-05-06...	in	?	0	235	?	don_aba...	just for fun, not...	26116	1172	242	6	Fal
8	PDI-P secara bulat d...	@Hotazilina	2023-05-06...	in	?	0	0	?	Nina Hotasi	Hilang hilang...	65	6	1	2	Fal
9	RT @wlpelmas: HJ...	@sofaven477	2023-05-06...	in	?	0	11	?	sofaven477	?	2332	3	1	0	Fal
10	RT @ceritudungimu...	@hobi613	2023-05-06...	in	?	0	143	?	alisa'	bts, genshin, an...	224223	1320	881	8	Fal
11	Tumpas habis KST P...	@YaniRS15	2023-05-06...	in	ID	0	0	?	Yani RS	Cinta tinggalka...	303	206	111	0	Fal
12	RT @muctar... BPP...	@nyoemhokgie	2023-05-06...	in	?	0	44	?	Ksatria Baj...	Oh Twitter, tunj...	33008	266	443	7	Fal
13	RT @ceritudungimu...	@fotnce	2023-05-06...	in	?	0	143	?	Memet Ke...	ic @@bl_blu...	9640	330	343	4	Fal
14	RT @wlpelmas: HJ...	@imacove1	2023-05-06...	in	?	0	11	?	moawee	?	2179	11	1	0	Fal
15	@igm_gm Dunggu Le...	@ari_jlg04	2023-05-06...	in	?	0	0	@igm_gm	Prof. (HCL)...	simply...	14652	191	99	1	Fal
16	RT @rustamaji: BPP...	@danangyulian...	2023-05-06...	in	?	0	344	?	Danang Y...	Tech Enthusiast...	68479	101	245	2	Fal
17	RT @wlpelmas: HJ...	@ciky42167257	2023-05-06...	in	?	0	11	?	ciky	?	2264	5	2	0	Fal
18	Super cakep guys ...	@dreimora	2023-05-06...	en	?	0	0	?	Raz	chaotic mind	179	21	1	0	Fal
19	Tumpah ruah padat...	@singghhmasu...	2023-05-06...	in	?	0	0	@singghh...	singgh m...	?	153	32	19	0	Fal
20	Disambut Antusias ...	@singghhmasu...	2023-05-06...	in	?	0	0	@singghh...	singgh m...	?	153	32	19	0	Fal
21	Bawa energi positif ...	@singghhmasu...	2023-05-06...	in	?	0	0	@singghh...	singgh m...	?	153	32	19	0	Fal
22	Berkunjung Ke Jawa...	@singghhmasu...	2023-05-06...	in	?	0	0	?	singgh m...	?	153	32	19	0	Fal
23	@Rigensih rigens si c...	@halleybieber	2023-05-06...	in	?	0	0	@Rigensih	mayki znn	this user is bad...	23846	237	119	2	Fal
24	@BPPSKesehatanRI ...	@fahruhudhas	2023-05-06...	in	?	0	0	@BPPSKese...	Fahru Hu...	Hai!	9443	157	373	2	Fal
25	RT @Restty_cayah...	@lukmamhsan	2023-05-06...	in	?	0	34	?	Budaya K...	?	103224	1190	964	6	Fal
26	RT @wlpelmas: HJ...	@yuan_sab6	2023-05-06...	in	?	0	11	?	paricola a...	*****aku sa...	147	68	10	0	Fal
27	RT @AindaFajriani...	@ArindaFajriani	2023-05-06...	ja	?	0	1	?	Arinda Faj...	Pegujung Skripsi	516	341	211	2	Fal
28	RT @viv_vistad Saan...	@EmiEvalin	2023-05-06...	in	?	0	21	?	Emi Evalin	?	67	1	417	0	Fal
29	RT @silasab6 Baka...	@ivan_pembur...	2023-05-06...	in	?	0	2	?	ivan pemb...	?	7208	1188	777	0	Fal
30	BPM&K... Tump...	@syaifulhuda...	2023-05-06...	in	?	0	106	?	?	?	37436	1000	710	0	Fal

Gambar. 5. Hasil pengambilan data twitter

Gambar 3 menunjukkan bahwa pengambilan data twitter dengan beberapa variabel yang ada, seperti *Content*, *Author*, *date*, dan variabel lain yang ada, namun data mentah ini masih terdapat simbol atau karakter pada tweet dari data yang diambil menggunakan API *Twitter*, sehingga perlu untuk adanya proses data agar data yang akan digunakan bersih dan hasil analisa dapat akurat tanpa adanya simbol dan karakter. Teks yang bersifat *noise*, simbol, ataupun tanda baca dihilangkan untuk mempersiapkan teks sebelum diolah karena dapat mempengaruhi nilai akurasi yang akan dihasilkan [12].

B. Preproses Data

Preprocessing text dari data yang ada merupakan tahapan untuk mempersiapkan data yang nantinya akan dianalisa, dengan tahapan sebagai berikut:



Gambar. 6. Alur *Preprocessing Data*

Berdasarkan gambar 3, *preprocessing text* dimulai dari data yang sudah didapatkan melalui proses *crawling*, lalu akan melalui proses *cleansing*, *Case folding*, *Tokenizing*, dan *Stemming*. Adapun hasil dari *text preprocessing* adalah file luaran excel.

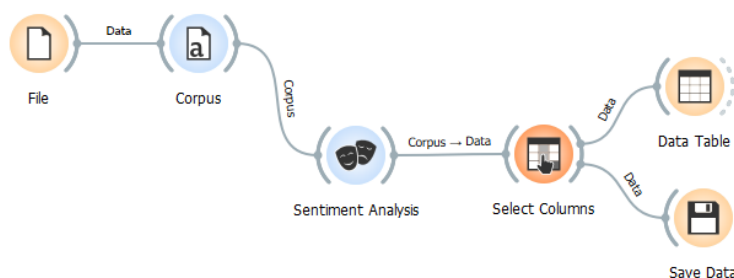
TABEL I
 CONTOH PERBEDAAN DATA SEBELUM DAN SESUDAH *PREPROCESSING TEXT*

Sebelum	Sesudah
@ngalammenfess Kalau masuk lewat igd bisa langsung claim bpjs nantinya (kalau punya bpjs) 😊.	ngalammenfess Kalau masuk lewat igd bisa langsung claim bpjs nantinya (kalau punya bpjs)
@hans_thetic Setau gua kisaran 50-150 klo di puskesmas. Bahkan pake bpjs gratis 😊	Hans thetic Setau gua kisaran 50 sampai 150 klo di puskesmas. Bahkan pake bpjs gratis
RT @RicKY_KCh: BPJS Warisan SBY .	RicKY KCh BPJS Warisan SBY .
Jangan sekali kali melupakan Sejarah , kecuali memang sudah Pikun ☐ https://t.co/SeXxy0YXCR	Jangan sekali kali melupakan Sejarah , kecuali memang sudah Pikun

Dengan hasil *preprocessing* tersebut, maka dapat dilakukan langkah selanjutnya dalam analisis data sentimen berdasarkan hal hal yang ada pada tahap preprocessing, yaitu *cleansing*, *Case folding*, *Tokenizing*, dan *Stemming*. Dengan menghasilkan kalimat yang sudah dibersihkan dan berbentuk kata perkata sehingga dalam suatu kalimat dapat diberi bobot berdasarkan kata positif, netral, dan negatif.

C. Analisis Sentimen Data

Dalam proses analisa sentiment ini, yang sebelumnya telah dilakukan proses data yang menjadikan data bersih dari simbol dan karakter yang ada, maka setelah itu dapat dilakukan penetapan sentimen terhadap *tweet* yang ada untuk menghitung sentimen apa yang dihasilkan oleh data tersebut.

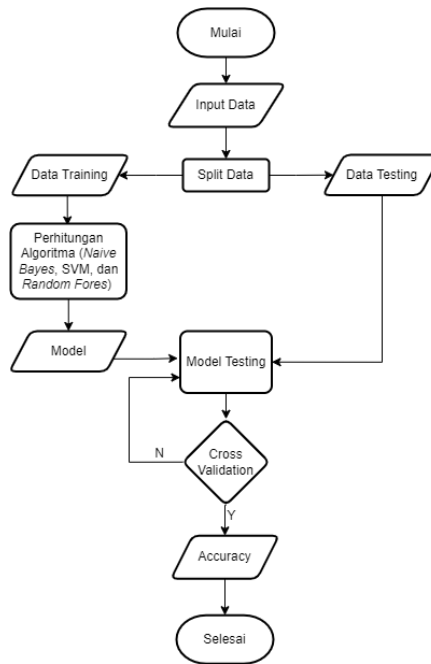


Gambar. 7. Proses Analisa Sentiment

Dengan data yang sudah mengalami *preprocessing data*, maka bisa disebutkan bahwa data yang akan dianalisa sudah layak untuk dilakukan Analisa sentimennya. Pada tahap *Sentiment Analysis*, pemilihan metode sentimen analisis *Multilingual Sentiment* dengan pemilihan bahasa Indonesia sebagai acuannya, dikarenakan sentimen terkait BPJS marak di wilayah Indonesia, serta memberi masukan teks berupa kata kata positif dan negatif menggunakan bahasa Indonesia terkait dengan data yang ada. Pada akhirnya melalui proses sentimen analisis ini akan menghasilkan luaran data berupa *numeric* atau angka negatif, netral, atau positif di setiap data *tweet*-nya. Dan pada tahap *Select Column* dapat memilih untuk kolom apa saja yang akan dipakai dan tidak, seperti contoh pada gambar 3 (tiga), terdapat banyak kolom atau variabel yang ada pada *dataset*, agar penelitian ini lebih fokus terhadap opini masyarakat maka dapat menggunakan fitur *Select Coloumn* ini untuk memilih kolom atau variable yang diperlukan, yaitu: Sentimen, *Author*, dan *Content*, dengan file *output* berupa excel.

D. Pemodelan

Setelah melakukan proses analisis sentimen maka data yang dihasilkan dapat untuk digunakan pendekatan algoritma yang ada untuk mendapatkan perhitungan dari yang sudah dilakukan. Dalam penelitian ini menggunakan pendekatan tiga algoritma, yaitu *Naïve Bayes*, *SVM*, dan *Random Forest*.



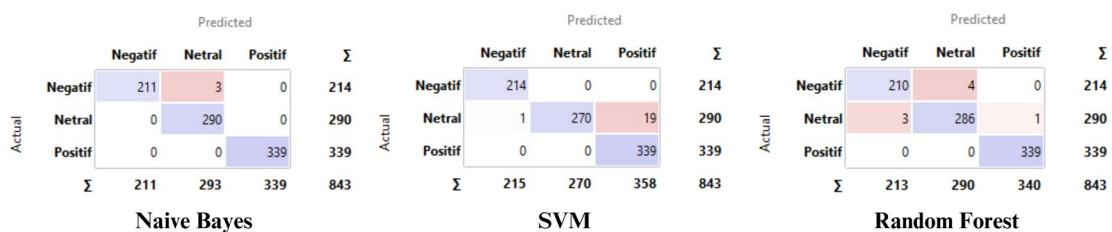
Gambar. 8. Proses Penggunaan Algoritma

Pada gambar 9 diatas merupakan pemodelan analisis sentimen menggunakan algoritma *Naive Bayes*, dimana dataset yang ada akan dilakukan split data menjadi data training dan data testing, dengan data training dilakukan perhitungan *Naive Bayes*, SVM, dan *Random Forest* yang ada, dan dari hasil perhitungan itu akan dimodelkan dan melakukan pengetesan dengan menggunakan *cross-validation* untuk melakukan pengulangan pengecekan perhitungan yang ada selama 10 kali, sehingga akan menghasilkan luaran berupa hasil perhitungan akurasi

E. Evaluasi

Data yang digunakan dalam penelitian ialah hasil dari proses sentimen analisis yang sudah dimodelkan dengan menggunakan beberapa metode yang menghasilkan akurasi pada setiap metodenya, dengan data yang diambil pada April 2023. Dalam mengolah data dilakukan analisis sentiment dari masyarakat Indonesia khususnya pada platform *Twitter* dengan tujuan mencari tahu pandangan masyarakat tentang program BPJS selama ini, apakah memberikan pandangan yang positif, netral, atau negatif.

Dengan hasil skor dan tes diatas maka didapatkan untuk precision setiap model atau metodenya hampir sama pada kisaran 99%. Dengan hasil nilai akurasi tersebut menandakan bahwa dengan menggunakan model *Naive Bayes*, SVM, dan *Random Forest* menjadi model yang efektif dan efisien untuk penelitian ini, namun hasil ini dapat dipengaruhi oleh karakteristik data yang diolah seberapa banyak dan bersih dari karakter selain teks, mengingat pada penelitian ini merupakan turunan dari *text mining*. Dan dengan hasil *test and score* tersebut dapat mengetahui bahwa data yang dipakai dan dilakukan sudah cukup akurat dalam proses data hingga analisis sentimen, dari hasil tersebut dapat divisualisasikan dengan diagram *Confusion Matrix*.



Gambar. 9. Hasil Visualisasi *Confusion Matrix*

Berdasarkan hasil *data training* diatas maka didapatkan untuk *Confusion Matrix* seperti gambar di atas, dengan hasil tersebut dapat menjadi acuan dalam pencarian hasil *Accuracy*, *Precision*, dan *Recall*. Dalam hal ini komponen yang ada pada tabel tersebut dapat dilakukan perhitungan, dengan hasil sebagai berikut:

TABEL II
HASIL PERHITUNGAN BERDASARKAN *CONFUSION MATRIX*

Metode	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
<i>Naïve Bayes</i>	99.6%	98.5%	98.9%
SVM	97.6%	93.1%	97.9%
<i>Random Forest</i>	98.5%	97.9%	98.3%

Menurut data yang ada pada Tabel 4 menunjukkan bahwa metode *Naïve Bayes*, SVM, dan *Random Forest* menghasilkan tingkat *Accuracy*, *Precision*, dan *Recall* yang cukup tinggi dalam proses perhitungan analisa dari data sentimen yang ada. Dengan menggunakan algoritma *Naïve Bayes* menghasilkan 339 sentimen positif, 290 sentimen netral, dan 211 sentimen negatif, dimana pada algoritma ini menjadikan sentimen terhadap program BPJS mayoritas beropini positif. Di sisi lain penggunaan algoritma SVM menghasilkan 339 sentimen positif, 270 sentimen netral, dan 214 sentimen negatif, yang dimana pada algoritma ini juga menghasilkan sentimen positif menjadi opini masyarakat Indonesia yang mendominasi. Selain itu pada algoritma *Random Forest* memiliki hasil 339 sentimen positif, 286 sentimen netral, dan 210 sentimen negatif, hal ini juga mendasari bahwa penggunaan algoritma *Random Forest* menghasilkan sentimen positif sebagai sentimen yang dominan. Hasil yang sangat baik untuk klasifikasi data opini masyarakat pada platform twitter, dengan nilai akurasi yang sangat tinggi dibandingkan dengan metode lainnya yaitu sebesar 99.6%. Pada penelitian ini menunjukkan bahwa metode *Naïve Bayes* dapat menjadi alternatif yang cocok untuk masalah klasifikasi, dan dengan metode ini, kemungkinan hasil akan bervariasi tergantung dari karakteristik yang digunakan. Hal ini akan menjadi keselarasan dengan penelitian sebelumnya yang dilakukan oleh Samsir dan tim (2021) bahwa ketepatan penggunaan metode *Naïve Bayes* yang sangat bagus dimana didapatkan tingkat akurasi dengan hasil 97,15% [11].

Sedikit berbeda dengan *Naïve Bayes*, terdapat perbedaan yang cukup tipis antara satu dengan yang lain, *Random Forest* dengan metode klasifikasi juga dapat menjawab seberapa efisien metode tersebut dapat dilakukan untuk penelitian ini. Meskipun pada penelitian sebelumnya terdapat perbedaan yang signifikan pada keakurasian klasifikasi yang dilakukan oleh Sipayung dan tim (2016) dengan hasil yang cukup baik yakni 77,8%, hal ini menandakan bahwa penggunaan metode *Random Forest* dalam penelitian ini memberikan hasil yang lebih baik jika melihat penelitian sebelumnya terkait pendeteksian sarkasme. Dari ketiga metode yang digunakan, metode *Support Vector Machine* atau SVM menjadi metode dengan hasil akurasi paling rendah jika dibandingkan dengan metode *Naïve Bayes* dan metode *Random Forest* pada penelitian ini. Pada penelitian yang dikaji oleh Setyawati dengan tim (2021) menunjukkan bahwa penggunaan *Random Forest* terhadap Analisa sentiment untuk program Kartu Prakerja mendapatkan hasil yang sangat bagus dengan tingkat akurasi mencapai 98,67%, hal ini sebanding dengan hasil yang dicapai pada penelitian ini terkait program BPJS [10]. Dengan ini menyatakan bahwa penggunaan metode *Naïve Bayes* mempunyai potensi untuk mencari hasil yang lebih baik dalam konsentrasi penelitian yang berbeda, meskipun metode SVM dan *Random Forest* memiliki keunggulan dan potensi menangani data yang mengandung *noise*.

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan dari penelitian yang telah dilakukan maka dapat disimpulkan bahwa dengan membandingkan metode *Naïve Bayes*, SVM, dan *Random Forest* dapat mengetahui sentimen masyarakat Indonesia yang terbaik. Dan dengan hasil sentimen analisis menggunakan data dari *twitter* ini dapat dilakukan dan mendapat hasil yang cukup variatif dikalangan masyarakat Indonesia dengan mayoritas memiliki opini atau pandangan yang positif terkait program BPJS ini dengan melihat *Confusion Matrix* yang ada, dengan metode SVM menjadi metode dengan hasil yang akurat dan efisien digunakan terkait dengan program Badan Penyelenggara Jaminan Sosial dengan nilai akurasi, presisi, dan *recall* mencapai 99.6%, dengan perbedaan yang tidak banyak terhadap metode *Naïve Bayes* dan *Random Forest* dibawah dari metode SVM. Sebagai saran pengembangan, dengan penelitian ini kiranya pada penelitian yang akan datang diharapkan dapat menggunakan pemasalahan baru dengan *dataset* yang lebih banyak dan tentunya dengan metode yang lebih bervariasi.

DAFTAR PUSTAKA

- [1] Suprpto and A. A. Malik, "Implementasi Kebijakan Diskresi Pada Pelayanan Kesehatan Badan Penyelenggara Jaminan Kesehatan (BPJS)" Jurnal Ilmiah Kesehatan Sandi Husada, Jun. 2019.
- [2] L. Linda, Y. Haskas, and E. Kadrianti, "Jurnal Ilmiah Kesehatan Diagnosis Volume 15 Nomor 2 Tahun 2020 eISSN : 2302-2531139 "Perbedaan Persepsi Pengguna Jasa Bpjs Dan Non Bpjs (Umum) Tentang Kualitas Pelayanan Keperawatandirsud Timika-Papua" Jurnal Ilmiah Kesehatan Diagnosis, 2020.
- [3] Ismatut Thobibah, Ernawaty, and Nyoman Anita Damayanti, "Analisis Faktor Penentu Kepesertaan BPJS," Jurnal Keperawatan Silampari 4, 2020.

- [4] R. Puspita and A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS," Jurnal Informatika Universitas Pamulang, 2020.
- [5] Y. Kurniawan, A. Pujiyanto, and S. Andayani, "Analisis Perbedaan Pelayanan Kesehatan Pada Pasien BPJS dan Pasien Umum (Non-BPJS)." [Online]. Available: <http://www.antarajatim.com/>,
- [6] Zamal Imron, "Positif dan Negatif BPJS," <https://www.kompasiana.com/zamaldikjemen/59db26d8bde57521d11c7742/positif-dan-negatif-bpjs>, Oct. 09, 2017.
- [7] Sopia Weni Anggriani, "Kualitas Pelayanan Bagi Peserta Bpjs Kesehatan Dan Non Bpjs Kesehatan" JISIP: Jurnal Ilmu Sosial dan Ilmu Politik, 2016.
- [8] Giovani AP, Ardiansyah, and Haryanti, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi " 2020.
- [9] M. K. Anam, M. I. Mahendra, W. Agustin, Rahmadden, and Nurjayadi, "Framework for Analyzing Netizen Opinions on BPJS Using Sentiment Analysis and Social Network Analysis (SNA) ," Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi, 2022.
- [10] Setyawati, N. Hendrastuty, A. R. Isnain, and A. Y. Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," Jurnal Informatika: Jurnal pengembangan IT, 2021.
- [11] Samsir, Ambiyar, U. Verawaedina, F. Edi, and R. Watrionthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," Jurnal Media Informatika Budidarma, 2021.
- [12] E. Indrayuni, A. Nurhadi, and D. A. Kristiyanti, "Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors Untuk Analisa Sentimen Aplikasi Halodoc," 2021.
- [13] I. K. Dharmendra, N. N. U. Januhari, I. P. Ramayasa, and I. M. A. W. Putra, "Uji Komparasi Sentiment Analysis Pada Opini Alumni Terhadap Perguruan Tinggi," Jurnal Teknik Informatika Unika ST. Thomas , 2022.
- [14] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," SMATIKA Jurnal, 2020.
- [15] A. S. Ritonga and E. S. Purwaningsih, "Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Kualitas Pengelasan Smaw (Shield Metal Arc Welding)," Jurnal Ilmiah Eduatic, 2018.
- [16] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," Indonesian Journal of Artificial Intelligence and Data Mining, 2018.
- [17] M. Y. Aldean, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac)," Journal of Informatics, Information System, Software Engineering and Applications, vol. Vol. 4 No.2, May 2022.
- [16] [18] D. Alita, and A. Rahman, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," Jurnal Komputasi, VOL. 8, NO. 2, 2020.