

Penerapan Algoritma *Support Vector Machine* Dalam Sentimen Analisis Terkait Kenaikan Tarif BPJS Kesehatan

Putu Mega Nirmala Dharmapatni¹, Ni Luh Putu Merawati*²

^{1,2}Jurusan Teknologi Informasi, Universitas Bumigora
meganirmala14@gmail.com¹, putu.mera@universitasbumigora.ac.id²

Abstrak

Jejaring sosial adalah alat komunikasi yang sangat populer di kalangan masyarakat salah satunya adalah jejaring sosial twitter. Jumlah tweets berkembang pesat karena kesederhanaan dan kemudahan penggunaannya menjadi alasan twitter lebih populer di kalangan masyarakat Indonesia dalam berkomunikasi. Twitter banyak digunakan sebagai media promosi produk, iklan, kampanye politik maupun sebagai sarana dalam menyampaikan pendapat terkait kritik, saran, isu-isu serta opini-opini yang sifatnya publik seperti pandangan netizen terhadap kebijakan baru pemerintah dan lain sebagainya. Salah satu kebijakan pemerintah yang sedang ramai diperbincangkan oleh netizen di jejaring sosial twitter adalah kenaikan tarif BPJS kesehatan. BPJS Kesehatan (Badan Jaminan Kesehatan dan Sosial) adalah Badan Usaha Milik Negara, besarnya jumlah pengguna BPJS menyebabkan BPJS harus menyediakan layanan umpan balik kepada pengguna untuk mengetahui respon pengguna terhadap layanan BPJS. Analisis sentimen atau disebut juga dengan penambangan pendapat adalah suatu bidang dalam menganalisis pendapat seseorang ataupun sekelompok orang mengenai suatu permasalahan yang tersedia untuk melihat respon masyarakat terhadap permasalahan tersebut. Metode SVM merupakan salah satu metode yang baik untuk digunakan dalam pengkalsifikasian data yang berjumlah banyak, terutama terkait dengan sentimen analisis. Pada penelitian ini akan dilakukan analisis sentimen masyarakat melalui media twitter terhadap kenaikan tarif BPJS Kesehatan menggunakan metode SVM untuk melihat respon masyarakat terhadap kebijakan tersebut. Berdasarkan hasil klasifikasi menggunakan algoritma SVM pada 671 sample data terdapat 271 data yang setuju dengan kebijakan yang dibuat serta 400 data yang tidak setuju terhadap kebijakan.

Kata Kunci : BPJS Kesehatan, Klasifikasi *Support Vector Machine*, Sentimen Analisis, Twitter

Abstract

Social networking is a communication tool that is very popular among the public, one of them is twitter. The number of tweets is growing rapidly because its simplicity and ease of use is the reason Twitter is more popular among Indonesians citizen. Twitter is widely used as a promotional medium of products, advertisements, political campaigns and as a means of conveying related criticisms, suggestions, issues and opinions of a public nature such as netizen views on the government's new policies and so on. One of the government's policies that is being discussed by netizens on social network twitter is the tariff on bpjs health. BPJS Kesehatan (Health and Social Security Agency) is a state-owned enterprise, the large number of BPJS users causes BPJS to provide feedback services to users to know the user's response to BPJS services. Sentiment analysis or so-called mining if there is anything in the analysis of Asia or sekeke people about a problem that is available to see the community's response to the problem. The SVM method is one of the good methods for use in a lot of calcification data, especially when it comes to analytical sentiment. In this study will be conducted public analysis through twitter media against BPJS health tariff using SVM method to see the public against the same policy. Based on the classification results using the SVM algorithm on 671 sample data, there are 271 data that agree with the policies made and 400 data that do not agree with the policy.

Keywords : BPJS Kesehatan, Classification, *Support Vector Machine*, Sentiment Analysis, Twitter

I. PENDAHULUAN

Perkembangan jejaring sosial sebagai alat komunikasi sangat populer di masyarakat sebagai salah satu media untuk berkomunikasi salah satunya adalah jejaring sosial Twitter. Berdasarkan data yang dirilis oleh Twitter Indonesia pada tahun 2016

menyebutkan bahwa Indonesia merupakan salah satu pengguna aktif yang produktif menulis tweets dengan presentase 77% pengguna, hal tersebut dapat di tunjukkan dari jumlah *tweets* yang dihasilkan sepanjang 2016 yang mencapai 4,1 miliar *tweets*. Jumlah *tweets* berkembang pesat karena

kesederhanaan dan kemudahan penggunaannya menjadi alasan Twitter lebih populer di kalangan masyarakat Indonesia dalam berkomunikasi. Twitter banyak digunakan sebagai media promosi produk, iklan, kampanye politik maupun sebagai sarana dalam menyampaikan pendapat terkait kritik, saran, isu-isu serta opini-opini yang sifatnya publik seperti pandangan *netizen* terhadap kebijakan baru pemerintah dan lain-lain sehingga informasi yang terkandung pada *tweets* ini sangat berharga sebagai data pendukung dalam pengambilan kebijakan[1][2].

BPJS Kesehatan adalah Badan Usaha Milik Negara yang khusus ditugaskan oleh pemerintah untuk mengelola kesehatan seluruh rakyat Indonesia yang telah terdaftar secara resmi. Besarnya jumlah pengguna BPJS menyebabkan BPJS harus menyediakan layanan umpan balik kepada pengguna untuk mengetahui respon pengguna terhadap layanan BPJS. Salah satu kebijakan pemerintah yang sedang ramai diperbincangan oleh *netizen* di jejaring sosial twitter adalah kenaikan harga BPJS kesehatan. Analisis sentimen atau disebut juga dengan penambangan pendapat adalah suatu bidang dalam menganalisis pendapat orang, evaluasi, penilaian, sikap dan emosi kepada suatu entitas seperti produk, layanan, organisasi, individu, masalah, acara dan topik. Fokus utama analisis adalah untuk mengekspresikan sentiment yang mencakup opini positif dan opini negatif[3][4].

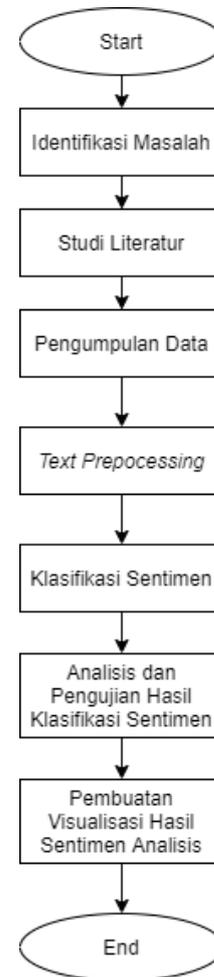
Metode SVM merupakan salah satu metode yang baik untuk digunakan dalam pengkalsifikasian data yang berjumlah banyak, terutama terkait dengan sentimen analisis[5], namun banyak penelitian yang dilakukan sebelumnya hanya membahas sentimen analisis dalam ruang lingkup yang terbatas dan jarang mengangkat tema mengenai kebijakan badan pemerintahan yang merupakan lembaga negara. Oleh karena itu pada penelitian ini akan dilakukan analisis sentimen masyarakat melalui media sosial *twitter* terhadap kenaikan tarif BPJS Kesehatan, yang diatur oleh Peraturan Presiden nomor 64 tahun 2020 tentang perubahan kedua atas Peraturan Presiden nomor 82 tahun 2018 menggunakan metode SVM untuk melihat respon masyarakat terhadap kebijakan tersebut, serta melihat tingkat kecocokan metode SVM dalam mengklasifikasikan kalimat ke dalam kelas positif dan negatif.

II. METODOLOGI

2.1. Tahapan Penelitian Secara Umum

Tahapan pertama dalam penelitian ini adalah melakukan indentifikasi masalah yakni terkait dengan kenaikan tarif BPJS Kesehatan berdasarkan

Peraturan Presiden Nomor 64 Tahun 2020 Tentang Perubahan Kedua Atas Peraturan Presiden Nomor 82 Tahun 2018.



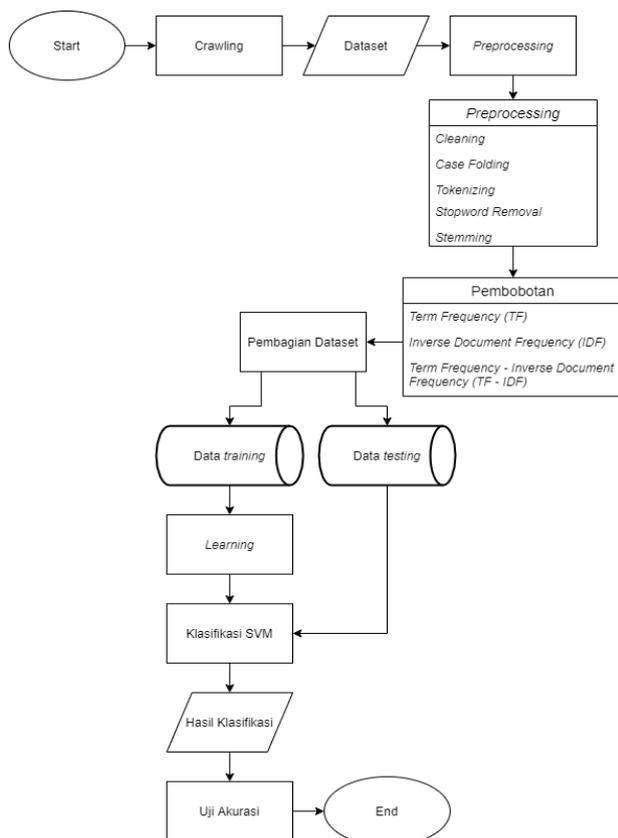
Gambar 1 Tahapan Penelitian Secara Umum

Studi literatur dilakukan untuk memperoleh gambaran, rujukan maupun referensi terkait topik permasalahan yang akan diangkat berdasarkan pada penelitian-penelitian lain yang telah dilakukan sebelumnya. Pengumpulan data atau *crawling* data merupakan proses penarikan data tweet dari media sosial twitter. Data yang digunakan dalam penelitian ini adalah data *primer*, yaitu data yang diambil langsung oleh penulis dari media sosial twitter. Data tersebut berupa *tweets* yang mencantumkan kata BPJS dan kenaikan tarif di dalamnya.

Data yang digunakan dalam penelitian ini berupa data *tweets* berisi kata sesuai topik yang sudah ditentukan untuk diteliti yaitu kenaikan tarif BPJS. *Tweets* data yang diambil dari bulan Januari 2020 sampai Juni 2020. Data *tweets* kemudian akan melalui proses *preprocessing* yaitu mengolah data mentah menjadi data yang lebih bersih untuk proses kalsifikasi. Sebelum proses klasifikasi dilakukan data hasil *preprocessing* akan dibagi kedalam data

training dan data *testing*. Data *training* adalah data yang akan digunakan untuk membentuk model pembelajaran dalam menentukan hasil sentimen. Setelah proses klasifikasi sentimen dilakukan selanjutnya adalah mengukur tingkat akurasi metode *support vector machine* dalam mengklasifikasikan sentimen kemudian hasil akhir dari proses klasifikasi adalah pembuatan visualisasi.

2.2 Deskripsi Sistem



Gambar 2 Sistem secara keseluruhan

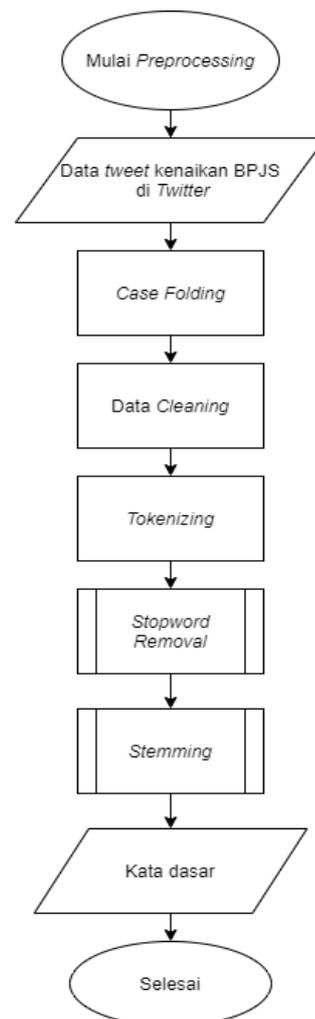
Data yang digunakan dalam penelitian ini berupa data berbentuk teks yang didapatkan dari tulisan masyarakat di media sosial Twitter. Data tersebut berupa *tweets* berisi kata sesuai topik yang sudah ditentukan untuk diteliti yaitu kenaikan tarif BPJS. *Tweets* data yang diambil dari bulan Januari 2020 sampai Juni 2020. Data *tweets* yang telah melalui proses *preprocessing*, sebagian akan digunakan sebagai data *training* yang merupakan model dalam menentukan hasil sentimen. Jenis data yang digunakan dalam penelitian ini adalah data *primer*, yaitu data yang diambil langsung oleh penulis dari media sosial twitter. Data tersebut berupa *tweets* yang mencantumkan kata BPJS dan kenaikan tarif di dalamnya.

Data *tweets* merupakan dokumen *tweets* berbahasa Indonesia yang didapat dari hasil *crawling*.

Crawling merupakan teknik mengumpulkan data pada sebuah *website* dengan memasukkan *Uniform Resource Locator (URL)*. URL ini menjadi acuan untuk mencari semua *hyperlink* yang ada pada *website*, kemudian dilakukan *indexing* untuk mencari kata dalam dokumen pada setiap *link* yang ada. Penerapannya *crawling* menggunakan *automation program* dan menggunakan *Application Programming Interface (API)* sebagai jalur komunikasi dalam mendapatkan data, dalam penelitian ini menggunakan *Application Programming Interface (API)* dari Twitter.

Data hasil *crawling* kemudian dibagi menjadi dua bagian, yaitu data *training* dan data *testing*. Data *training* merupakan data yang sudah diberi pelabelan sebelumnya, sedangkan data *testing* merupakan data yang akan diuji menggunakan model yang sudah dibuat. Hasil dari pengujian data *testing* tersebut merupakan hasil prediksi dari metode yang digunakan.

2.3 Preprocessing



Gambar 3 Preprocessing

Sentimen analisis merupakan salah satu sub ilmu dari *text mining*, *natural language program*, dan *artificial intelligence*. Proses pada analisis sentiment yaitu untuk memahami, mengekstrak, dan mengolah data teks secara otomatis sehingga menjadi suatu informasi yang bermanfaat[6]. Analisis sentimen merupakan bidang ilmu yang menganalisis pendapat, sikap, evaluasi, dan penilaian terhadap suatu peristiwa, topik, organisasi, maupun perseorangan[7].

Sentimen analisis merupakan salah satu metode untuk menganalisis sebagian data untuk mengetahui emosi manusia. Analisis sentimen dapat dikategorikan kedalam tiga *task*, yaitu *informative text detection*, *information extraction* dan *sentiment interestingness classification* (polarity identification, *emotional*). *Sentiment classification* (positif atau negatif) digunakan untuk memprediksi *sentiment polarity* berdasarkan data sentimen dari pengguna[8].

Preprocessing sangat menentukan hasil sentimen dengan metode SVM. *Preprocessing* merupakan tahap awal dari *text mining* untuk mengubah data sesuai dengan *format* yang dibutuhkan. Proses ini dilakukan untuk menggali, mengolah dan mengatur informasi dan untuk menganalisis hubungan tekstual dari data terstruktur dan data tidak terstruktur[7].

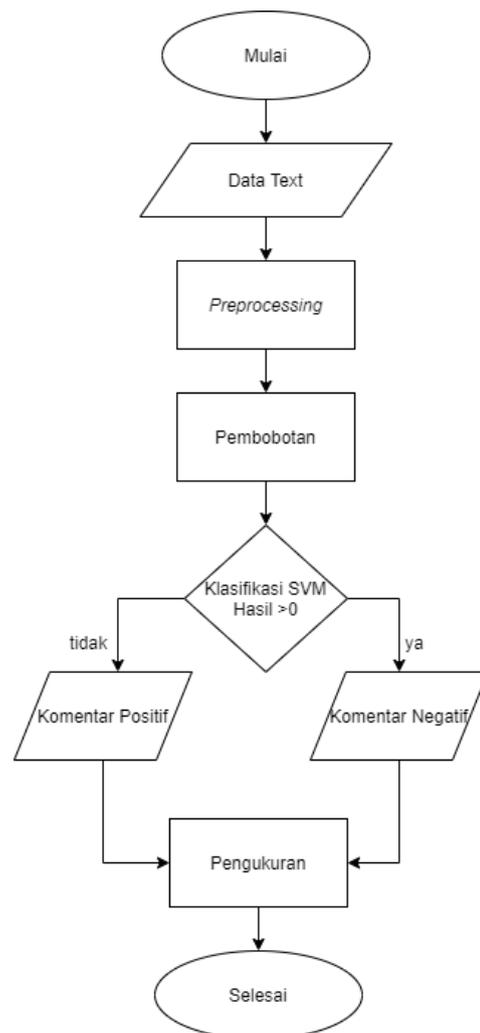
Tahap *preprocessing* terdiri atas beberapa proses antara lain adalah *case folding*, *data cleaning*, *tokenizing*, *stopword removal*, dan *stemming*. Data komentar *Twitter* mengenai kenaikan BPJS yang telah *dicrawling* serta disimpan dalam bentuk *.csv* selanjutnya akan dibersihkan agar data dapat diklasifikasi lebih lanjut menggunakan metode *Support Vector Machine*. Tahap awal dalam *preprocessing* data setelah dilakukan *crawling* adalah *cleansing*. *Cleansing* merupakan proses pembersihan data yang terbagi menjadi dua tahap yaitu *case folding* dan *data cleaning* atau *symbol removal*. *Cleansing* bertujuan agar data yang akan diolah menjadi lebih bersih dan tidak banyak mengandung *noise* yang dapat mempengaruhi hasil klasifikasi. Tahapan selanjutnya setelah proses *cleaning* adalah melakukan *tokenizing*.

Tokenizing bertujuan untuk mengumpulkan jumlah kata yang menjadi penyusun kalimat di dalam dataset. Data yang dimaksud merupakan kata tunggal, artinya jika ada 2 kata atau lebih di dalam *dataset*, maka hanya satu yang akan digunakan. Tahapan selanjutnya adalah melakukan *stopword removal*. *Stopword removal* adalah proses penyaringan untuk menghilangkan kata yang tidak relevan pada hasil *tokenizing* dengan cara membandingkannya dengan (*stopword list*) yang ada. Contoh dari *stopword* misalnya, kata sambung, artikel, dan preposisi.

Stemming merupakan proses untuk mencari *stem* (kata dasar) dari kata hasil *stopword removal*

(*filtering*). Terdapat dua aturan dalam melakukan *stemming* yaitu dengan pendekatan kamus dan pendekatan aturan[9]. *Stemming* yang digunakan pada skripsi ini adalah pendekatan kamus, dimana menggunakan *library* sastrawi. Data yang telah melalui tahapan *preprocessing* selanjutnya dapat diolah lebih lanjut yakni pemberian bobot pada setiap data *tweet* yang telah dibersihkan.

2.4 Klasifikasi



Gambar 4 Klasifikasi SVM

Pada Gambar 4 dijelaskan data berupa kata-kata tweets yang sudah melalui proses *preprocessing* kemudian dihitung pembobotannya, setelah pembobotan diklasifikasi, apabila kata atau term tersebut bernilai lebih besar sama dengan 0 maka kata tersebut akan dikategorikan sebagai komentar positif, apabila kata atau term bernilai lebih kecil dari 0 maka kata tersebut akan dikategorikan sebagai komentar negatif.

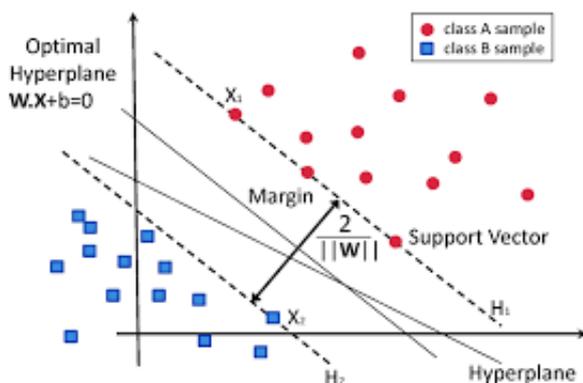
SVM dalam proses klasifikasi bertujuan untuk menemukan garis yang terbaik (*hyperplane*) untuk membagi dua kelas, kemudian data diklasifikasi

berdasarkan pada sisi mana garis tersebut muncul. SVM merupakan metode klasifikasi pada metode *machine learning (supervised learning)* yang memprediksi kelas berdasarkan pola dari hasil proses *training* yang diciptakan oleh Vladimir Vapnik.

Klasifikasi dilakukan dengan garis pembatas (*hyperplane*) yang memisahkan antara kelas opini positif dan opini negatif. Secara intuitif, suatu garis pembatas yang baik adalah yang memiliki jarak terbesar ke titik data pelatihan terdekat dari setiap kelas, karena pada umumnya semakin besar margin, semakin rendah *error* generalisasi dari pemilah. *Margin* adalah jarak dari suatu titik vektor di suatu kelas terhadap *hyperplane*[8].

Klasifikasi menggunakan *Support Vector Machine* akan dimulai dengan mengubah teks menjadi data vektor. Data vektor tersebut mempunyai dua dimensi yaitu (*word id*) dan bobot. SVM dalam klasifikasi memberikan setiap kata di dalam dokumen data *tweets* dimensi dan sebuah bobot berdasarkan seberapa penting kata tersebut di dalam dokumen. SVM menentukan garis terbaik yang memisahkan dua kelas di dalam sebuah proses kalsifikasi[10].

Secara sederhana pada konsepnya cara kerja SVM (*Support Vector Machine*) adalah mencari *Hyperplane* terbaik, yang berfungsi sebagai pemisah dua buah kelas pada *input space*. Dua kelas, +1 dan -1, beserta masing-masing *pattern* terlihat pada gambar dibawah:



Gambar 5 Grafik *Support Vector Machine*

Pada Gambar 5 dijelaskan dalam proses mengkalsifikasi untuk mendapat hasil yang baik, *hyperplane* digunakan sebagai pemisah dua kelas yang berbeda, dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya[9]. *Margin* adalah jarak antara *hyperplane* dengan *pattern* terdekat dari masing-masing kelas. *Pattern* terdekat dengan *hyperplane* disebut dengan *support vector* pada gambar diatas ditunjukkan dengan X_1 dan X_2 atau *point* yang berada pada garis putus-putus.

Berdasarkan Gambar 5, bidang pemisah dapat dirumuskan dengan:

$$m = \frac{1-b-(-1-b)}{|w|} = \frac{2}{|w|} \quad (1)$$

- m = jarak antara dua bidang
- w = bidang normal
- b = posisi relative terhadap origin

jarak garis dirumuskan $wx + b = c$ dan ke *origin* adalah $\frac{(c-b)}{|w|}$

Margin m dimaksimalkan dengan memenuhi konstrain 2 bidang pembatas pada *support vector*. Bidang pembatas kelas pertama membatasi kelas pertama yang bernilai + sedangkan bidang pembatas kelas kedua membatasi kelas kedua yang bernilai - sehingga diperoleh persamaan:

$$x_i \cdot w + b \geq +1 \text{ for } y_i = +1 \quad (2)$$

$$x_i \cdot w + b \leq -1 \text{ for } y_i = -1 \quad (3)$$

Untuk mencari nilai margin terbesar pada *hyperplane* atau garis pemisah antar dua buah kelas, dapat dirumuskan kedalam optimasi konstrain sebagai berikut:

$$\min = \frac{1}{2} |w|^2 \quad (4)$$

$$s. t \ y_i(x_i \cdot w + b) - 1 \geq 0 \quad (5)$$

Dengan lebih mudah untuk menyelesaikan permasalahan optimasi konstrain dalam menyelesaikannya dirubah kedalam formula lagrangian yang menggunakan *lagrange multiplier* yang diubah menjadi:

$$\min_{w,b}^{Lp}(w, b, a) = (w, b, a) \quad (6)$$

$$= \frac{1}{2} |w|^2 - \sum_{i=1}^n \alpha_i y_i(x_i \cdot w + b) + \sum_{i=1}^n \alpha_i$$

Permasalahan pencarian bidang pemisah terbaik dapat dirumuskan sebagai berikut:

$$\max_{\alpha} L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j \quad (7)$$

$$s. t \ \sum_{i=1}^n \alpha_i y_i = 0 \ \alpha_i \geq 0 \quad (8)$$

Untuk mendapatkan nilai α_i yang digunakan untuk menemukan w, maka terdapat α_i untuk setiap data yang digunakan untuk pelatihan. Data untuk

pelatihan yang mempunyai nilai $\alpha_i > 0$ disebut *support vector*, sedangkan yang memiliki nilai $\alpha_i = 0$ adalah sisanya.

Formula pencarian bidang pemisah terbaik atau *hyperplane* terbaik adalah permasalahan *quadratic programming*, sehingga nilai maksimum *global* dari α_i akan selalu dapat ditemukan setelah solusi permasalahan *quadratic programming* ditemukan (nilai α_i), maka kelas dari data pengujian x dapat ditemukan berdasarkan nilai dari fungsi keputusan:

$$f(x_d) = \sum_{i=1}^{ns} \alpha_i y_i x_i x_d + b \quad (10)$$

x_i = support vector

ns = jumlah support vector

x_d = data yang akan dikalsifikasikan

2.5 Analisis dan Pengujian Hasil Klasifikasi

Akurasi klasifikasi merupakan ukuran ketepatan klasifikasi yang menunjukkan performansi teknik klasifikasi secara keseluruhan[6]. Semakin tinggi akurasi klasifikasi berarti performansi teknik klasifikasi juga semakin baik. Kinerja metode klasifikasi dalam melakukan pengklasifikasian menggambarkan seberapa baik metode tersebut dalam mengklasifikasikan data.

Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya[6]. Pada pengukuran kinerja menggunakan *confusion matrix* terdapat 4 istilah yang merepresentasikan hasil proses klasifikasi, seperti yang terlihat pada Tabel 1:

Tabel 1 Confussion Matrix

Aktual	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan :

TP : *True Positive* (Jumlah prediksi benar pada kelas positif)

FP : *False Positive* (Jumlah prediksi salah pada kelas positif)

FN : *False Negative* (Jumlah prediksi salah pada kelas negatif)

TN : *True Negative* (Jumlah prediksi benar pada kelas negatif).

Berdasarkan nilai dari keempat klasifikasi tersebut, maka akan diperoleh nilai akurasi, presisi, dan *recall*. Nilai akurasi adalah nilai yang menggambarkan seberapa akurat metode yang digunakan dalam melakukan klasifikasi dengan benar dari keseluruhan data yang ada. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang dikalsifikasikan positif. Nilai *recall* menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar.

$$\text{Akurasi} = \frac{TN+TP}{TN+TP+FN+FP} \times 100\% \quad (11)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\% \quad (12)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (13)$$

2.6 Pembuatan Visualisasi Hasil Sentimen Analisis

Visualisasi hasil klasifikasi sentimen pada penelitian ini menggunakan pie diagram. Pie diagram adalah gaya grafik sederhana yang berguna untuk menggambarkan bagian-bagian dari keseluruhan data. Pie diagram adalah bagan berdasarkan data numerik untuk menunjukkan proporsi dalam persentase. Pie diagram pada penelitian ini digunakan untuk mempermudah melihat proporsi persentase jumlah dataset yang terklasifikasi kedalam komentar positif dan komentar negatif.

III. HASIL DAN PEMBAHASAN

3.1 Hasil Crawling Data

Pada proses *crawling* data *twitter* pada penelitian ini menggunakan bantuan aplikasi *google spreadsheet* dan serta *tools add-ons* yaitu *twitter archiver*. *Add-Ons* adalah salah satu komponen opsional yang terdapat didalam menu *google spreadsheet*, *add-ons* biasanya digunakan untuk menambah program atau fungsi-fungsi tambahan lainnya pada file Excel, Word, ataupun file lainnya yang dibuat menggunakan *google spreadsheet*. *Add-Ons* yang digunakan untuk *crawling* data *twitter* adalah *Tweet Archiver*.

Twitter archiver merupakan salah satu *add-on* dari Google Chrome yang bisa digunakan untuk mengumpulkan data *tweet* dengan *keyword*, *hashtag*, atau *mention* tertentu yang kemudian akan disimpan kedalam Google Spreadsheet. Secara berkala *twitter archiver* akan mengumpulkan *tweet* baru sesuai kriteria pencarian yang sudah disetting dan *google*

spreadsheet akan otomatis mengupdate *tweet-tweet* baru yang ada jika *tweet* tersebut sesuai dengan *rule* yang telah dibuat sebelumnya.

Rule adalah kriteria yang akan digunakan pada proses *crawling* data. Data yang *dicrawling* akan sesuai dengan *rule* yang telah di tentukan kriteria pencarian (*search rule*) dapat dilakukan dari yang sangat sederhana sampai yang kompleks. Untuk mendapatkan data mengenai kenaikan tarif BPJS Kesehatan dapat menginputkan filter terkait *keyword* kenaikan BPJS.

#	Time	Screen Name	Full Name	Tweet Text	Tweet ID	Lang	Mode	Location
1	2020/02/08 15:02	@yudi27	yudiThawwa	Bayer ngungap ngungap abar tidak bagean PMK. Jikalau Rongga	12776227048840793			
2	2020/02/08 15:02	@BudiArdana	Budi Ardana	Di koridor kesehatan seperti ini. Penderitaan COVID-19	1281220762046102			
3	2020/02/08 15:01	@BudiArdana	Budi Ardana	Yang penting kesehatan. Bekas hasil juga, jadi PMK yang	1281220440000000			
4	2020/02/08 14:58	@BudiArdana	Budi Ardana	Si baru: bpjs kesehatan naik ada bagean dan ditingkatkan	1281118861911287			
5	2020/02/08 14:52	@BudiArdana	Budi Ardana	bpjs naik, kesehatan kesehatan seperti ini	1281099200000000			
6	2020/02/08 14:52	@BudiArdana	Budi Ardana	Keputusan bpjs naik, bpjs naik, bpjs naik	1281099200000000			
7	2020/02/08 14:52	@BudiArdana	Budi Ardana	Keputusan bpjs naik, bpjs naik, bpjs naik	1281099200000000			
8	2020/02/08 14:52	@BudiArdana	Budi Ardana	Keputusan bpjs naik, bpjs naik, bpjs naik	1281099200000000			
9	2020/02/08 14:52	@BudiArdana	Budi Ardana	Keputusan bpjs naik, bpjs naik, bpjs naik	1281099200000000			
10	2020/02/08 14:52	@BudiArdana	Budi Ardana	Keputusan bpjs naik, bpjs naik, bpjs naik	1281099200000000			
11	2020/02/08 14:52	@BudiArdana	Budi Ardana	Keputusan bpjs naik, bpjs naik, bpjs naik	1281099200000000			
12	2020/02/08 14:52	@BudiArdana	Budi Ardana	Keputusan bpjs naik, bpjs naik, bpjs naik	1281099200000000			
13	2020/02/08 14:52	@BudiArdana	Budi Ardana	Keputusan bpjs naik, bpjs naik, bpjs naik	1281099200000000			
14	2020/02/08 14:52	@BudiArdana	Budi Ardana	Keputusan bpjs naik, bpjs naik, bpjs naik	1281099200000000			
15	2020/02/08 14:52	@BudiArdana	Budi Ardana	Keputusan bpjs naik, bpjs naik, bpjs naik	1281099200000000			
16	2020/02/08 14:52	@BudiArdana	Budi Ardana	Keputusan bpjs naik, bpjs naik, bpjs naik	1281099200000000			
17	2020/02/08 14:52	@BudiArdana	Budi Ardana	Keputusan bpjs naik, bpjs naik, bpjs naik	1281099200000000			

Gambar 6 Hasil Dashboard Pada Indeks KAMI

#	Location	Retweets	Replies	App	Follows	Influ	Listed	Verified	User Since	Location	Bio	Website
1		0	0	0	0	0	0	0	10/2020	North Dakota	@PROVIDENCE	
2		0	0	0	0	0	0	0	4/2011	Bandung, Indonesia	Janis Y. Widiarta	
3		0	0	0	0	0	0	0	12/2019	02000-Jakarta, Indonesia	@yudi27	
4		0	0	0	0	0	0	0	4/2020	Jakarta, Jakarta	Indah Rizki	
5		0	0	0	0	0	0	0	3/2020	Karangany, Indonesia	Budi Ardana	
6		0	0	0	0	0	0	0	3/2017	Jakarta, Indonesia	Budi Ardana	
7		0	0	0	0	0	0	0	4/2020	Jakarta, Jakarta	Budi Ardana	
8		0	0	0	0	0	0	0	4/2020	Jakarta, Jakarta	Budi Ardana	
9		0	0	0	0	0	0	0	4/2020	Jakarta, Jakarta	Budi Ardana	
10		4	14	0	0	0	0	0	1/2020	Jakarta, Jakarta	Budi Ardana	
11		0	0	0	0	0	0	0	1/2020	Jakarta, Jakarta	Budi Ardana	
12		0	0	0	0	0	0	0	10/2020	Jakarta, Jakarta	Budi Ardana	
13		0	0	0	0	0	0	0	3/2020	Jakarta, Jakarta	Budi Ardana	
14		0	0	0	0	0	0	0	3/2020	Jakarta, Jakarta	Budi Ardana	
15		0	0	0	0	0	0	0	2/2020	Jakarta, Jakarta	Budi Ardana	
16		0	0	0	0	0	0	0	3/2020	Jakarta, Jakarta	Budi Ardana	
17		0	0	0	0	0	0	0	8/2018	Indonesia	shut up go home	

Gambar 7 Hasil Dashboard Pada Indeks KAMI

Gambar 6 dan Gambar 7 merupakan hasil dari proses *crawling* data menggunakan *twitter archiver*. Data hasil *crawling* akan sesuai dengan *rule* yang telah ditentukan sebelumnya. Twitter archiver dapat memperoleh data *tweet* dengan informasi yang sangat lengkap. Data hasil *crawling* dapat disimpan dalam berbagai jenis format salah satunya adalah CSV (*Comma Separated Values*).

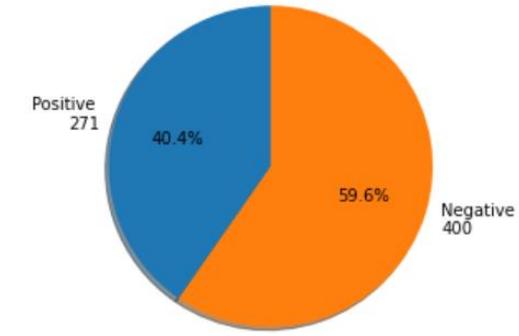
3.2 Hasil Klasifikasi

Proses pembentukan data *training* pada penelitian ini menggunakan metode *cross-validation*, dengan jumlah validasi dari nilai x yang di uji cobakan sebanyak 1-10 kali. Data *testing* pada penelitian ini digunakan sebanyak 20% dan data *training* sebanyak 80% dari total keseluruhan *sample* data yang ada. Gambar 8 dan Gambar 9 menunjukkan hasil klasifikasi yang dilakukan oleh SVM. Berdasarkan diagram pada Gambar 9 terlihat bahwa dari total 671 data terdapat 271 data yang termasuk

kategori positif serta 400 data termasuk kedalam kategori sentimen negatif.

	Tweet Text	Sentimen
2		
3	Bpjs naik, diem. Listrik naik, diem. Bbm naik, diem. Pancasila mau dirubah, juga diem. Eh giliran benderanya dibakar malah demo 🤔🤔	negatif
4	Iuran BPJS Kesehatan Akan Naik Lagi Bulan Depan	positif
5	Sebentar-sebentar. Bpjs naik lagi per 1 Juli Gol 1 150k Gol 2 100k Gol 3 42k Tapi Gol 3 tetap membayar seperti biasa yaitu 25.5k dan sisanya 16.5k dibayar oleh pemerintah. Terus per Januari 2021 Gol 3 bayarnya 35k dan 7k oleh pemerintah. Apasih. Gajelas. Ngapain ngasih bantuan.	negatif
6	#TurunTurunTurunJokowi #makzulkanjkw Rezim anti kritik, Rezim zolim, mencekik ekonomi rakyat. BPJS naik, BBM Gak turun disaat minyak dunia turun, anggaran covid 900T gak dirasakan masyarakat bawah, sembako gak berkualitas (merk sembaknya gak pernah terdengar), UUD PIP Komunis.	negatif
7	Listrik NAIK diam! BPJS naik diam! Pertamina ga turunkan BBM saat harganya turun, diam! Begitu emaknya disentil, bantengnya disenggol, turun jalan! Emang klo dah gaya preman susah sih...Pengennya rame2, mo pawai motor kalee...	negatif
8	Siap-Siap Iuran BPJS Kesehatan Akan Naik Lagi. Ini daftar Harga Per Kelasnya	positif
9	Sekrng apa" serba "Naik Naik Naik" rakyat sudah terbebani kebijakan ambradul rezim, sekrng rakyat butuh "Turun Turun Turun" bukan sembako, listrik bpjs, bbm, dll. Tapi presiden @Jokowi Turun3	negatif
10	#TurunTurunTurunJokowi Mba @LunaMaya26 Sedih itu bukan lari diudara yg anda mksd. Sedih itu melihat TKA kerja rakyat nganggur, ulama & agama dihina, perkosaan pencabulan anak, hukum tdk tegak, korupsi merajalela, bhn pangan bbm bpjs listrik naik, dan zina/sex bebas dipertontonkan di media internet.	negatif

Gambar 8 Hasil Klasifikasi Sentimen



Gambar 9 Diagram Hasil Klasifikasi

3.3. Hasil Pengujian

Pada Gambar 10 menunjukkan hasil perhitungan dari *confusion matrix*. Pengujian dilakukan untuk mengetahui tingkat akurasi klasifikasi hasil sentimen analisis menggunakan metode SVM pada kasus kenaikan tarif BPJS Kesehatan.

	precision	recall	f1-score	support
0	0.90	0.96	0.93	376
1	0.94	0.87	0.90	295
accuracy			0.92	671

accuracy_score = 0.9195230998509687
Number of Data Frame Test : 671

Gambar 10 Hasil Tingkat Akurasi SVM

Berdasarkan nilai dari keempat klasifikasi tersebut, maka akan diperoleh nilai akurasi, presisi, dan *recall*. Nilai akurasi adalah nilai yang menggambarkan seberapa akurat metode yang digunakan dalam melakukan klasifikasi dengan benar dari keseluruhan data yang ada. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasikan positif. Nilai *recall* menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar.

Tabel 2 *Confusion Matrix*

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	256	361
Negatif	15	39

Tabel 2 menunjukkan hasil klasifikasi sentimen menggunakan metode SVM yang telah dikelompokkan menggunakan tabel *confusion matrix*.

IV. KESIMPULAN

Hasil penelitian yang dilakukan menunjukkan metode SVM memiliki tingkat akurasi atau tingkat kedekatan antara nilai prediksi dari sistem dengan nilai aktual yang mencapai 92% pada proses pengujian menggunakan *confusion matrix*. Hasil *precision* atau tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem pada sentimen positif sebesar 94%, *recall* yang merupakan nilai rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif sebesar 87% dan hasil *precision* pada sentimen negatif sebesar 90%, serta *recall* pada sentimen negatif sebesar 96%.

Berdasarkan hasil klasifikasi menggunakan algoritma SVM pada 671 sample data terdapat 271 data yang setuju dengan kebijakan yang dibuat serta 400 data yang tidak setuju terhadap kebijakan. Hal tersebut menunjukkan bahwa sentimen masyarakat terkait dengan kebijakan BPJS Kesehatan yang menaikkan tarif pembayaran cenderung bernilai negatif atau tidak setuju.

REFERENSI

[1] K. M. Carley, M. Malik, M. Kowalchuck, J. Pfeffer, and P. Landwehr, "Twitter Usage in Indonesia," *SSRN Electron. J.*, no. August 2016, 2018,
[2] K. M. Carley, M. Malik, P. M. Landwehr, J. Pfeffer, and M. Kowalchuck, "Crowd sourcing disaster management: The complex

nature of Twitter usage in Padang Indonesia," *Saf. Sci.*, vol. 90, pp. 48–61, 2016
[3] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," *Proc. 7th Int. Conf. Lang. Resour. Eval. Lr. 2010*, pp. 1320–1326, 2010
[4] A. Z. Amrullah, A. S. Anas, M. Adrian, J. Hidayat, F. Teknik, and U. Bumigora, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *J. BITE J. Bumigora Inf. Technol.*, vol. 2, no. 1, pp. 40–44, 2020
[5] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1725–1732, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628>.
[6] A. D. Hartanto and N. K. Fitriyani, "Analisis Sentimen Terhadap Tokoh Publik Menggunakan Support Vector Machine," *MEANS (Media Inf. Anal. dan Sist.*, vol. 5, no. 1, pp. 8–12, 2020, [Online]. Available: http://ejournal.ust.ac.id/index.php/Jurnal_Means/.
[7] Y. X. Chu, X. G. Liu, and C. H. Gao, "Application of Support Vector Machine in Bioinformatics," *Proc. 2011 Chinese Control Decis. Conf. CCDC 2011*, pp. 842–847, 2011
[8] R. Valentini, P. Siwabessy, A. Herdiani, and A. Romadhony, "Analisis Sentimen Masyarakat Terhadap Hasil Kerja Petahana Dalam Kaitan Dengan Pemilihan Presiden tahun 2019 Pada Sosial Media Twitter Menggunakan Support Vector Machine (SVM)," vol. 6, no. 2, pp. 8625–8636, 2019.
[9] M. S. Utomo, "Implementasi Stemmer Tala pada Aplikasi Berbasis Web," *J. Teknol. Inf. Din.*, vol. 18, no. 1, pp. 41–45, 2013.
[10] Q. Han, J. Guo, and H. Schütze, "CodeX: Combining an SVM classifier and character N-gram language models for sentiment analysis on twitter text," **SEM 2013 - 2nd Jt. Conf. Lex. Comput. Semant.*, vol. 2, no. SemEval, pp. 520–524, 2013.