

Analisis Sentimen Terhadap Kenaikan Iuran Bpjs Kesehatan Pada *Twitter* Menggunakan *Naive Bayes Classifier*

Putri Dewi Wulandari ^{1✉}, Ultach Enri ², Aji Primajaya ³

¹⁾²⁾³⁾Program Studi Teknik Informatika, Universitas Singaperbangsa Karawang, Indonesia

¹⁾putri.16166@student.unsika.ac.id

²⁾ultach@staff.unsika.ac.id

³⁾aji.primajaya@staff.unsika.ac.id

Abstract – Health service is one of the fundamental rights of the community whose provision must be carried out by the Government as mandated in the 1945 Constitution Article 28 H Paragraph (1). According to the Presidium of the Indonesian Student Press Association Activist Alumni Forum, the National Health Insurance Program whose implementation is entrusted to the Health BPJS is far from the meaning of justice. One of the problems faced by the Health BPJS is the large number of participants who are in arrears in making payments, causing a huge deficit in the Health BPJS and causing the Government to officially raise BPJS Health contributions according to the Minister of Finance's proposal Sri Mulyani. The number of public complaints about government policy issues in raising BPJS Health fees, so this study was conducted with the aim of finding conclusions from public perceptions using the Naive Bayes Classifier algorithm with the TF-IDF selection feature. The dataset used as many as 356 data divided into two classes, namely positive and negative classes that have been through the selection phase by Indonesian Language Experts. The classification process produces the best accuracy results with a split ratio of 60:40 that is equal to 92.96% using the TF-IDF selection feature.

Keywords – Sentiment, Naive Bayes Classifier, BPJS Health, Text Mining, TF-IDF

Intisari – Pelayanan kesehatan merupakan salah satu hak mendasar masyarakat yang penyediaannya wajib diselenggarakan oleh Pemerintah sebagaimana telah diamanatkan dalam Undang-Undang Dasar 1945 Pasal 28 H Ayat (1). Menurut Presidium Forum Alumni Aktivist Perhimpunan Pers Mahasiswa Indonesia, Program Jaminan Kesehatan Nasional yang pelaksanaannya dipercayakan pada BPJS Kesehatan masih jauh dari makna keadilan. Salah satu persoalan yang dihadapi BPJS Kesehatan yaitu banyaknya peserta yang menunggak dalam melakukan pembayaran sehingga menimbulkan defisit yang sangat besar pada BPJS Kesehatan dan menyebabkan Pemerintah resmi menaikkan iuran BPJS Kesehatan sesuai usulan Menteri Keuangan Sri Mulyani. Banyaknya keluhan masyarakat terhadap isu kebijakan Pemerintah dalam menaikkan iuran BPJS

Kesehatan, maka dilakukan penelitian ini dengan tujuan untuk menemukan kesimpulan dari persepsi masyarakat menggunakan algoritma *Naive Bayes Classifier* dengan fitur seleksi TF-IDF. *Dataset* yang digunakan sebanyak 356 data yang terbagi kedalam dua kelas yaitu kelas positif dan negatif yang telah melalui tahap seleksi oleh Ahli Bahasa Indonesia. Proses klasifikasi menghasilkan hasil akurasi terbaik dengan *ratio split* 60:40 yaitu sebesar 92,96% dengan menggunakan fitur seleksi TF-IDF.

Kata Kunci – Sentimen, Naive Bayes Classifier, BPJS Kesehatan, Text Mining, TF-IDF

I. PENDAHULUAN

Pelayanan kesehatan merupakan salah satu hak mendasar masyarakat yang penyediaannya wajib diselenggarakan oleh Pemerintah sebagaimana telah diamanatkan dalam Undang-Undang Dasar 1945 Pasal 28 H Ayat (1) yang berbunyi “Setiap orang berhak hidup sejahtera lahir dan batin, bertempat tinggal, dan mendapatkan lingkungan hidup yang baik dan sehat serta berhak memperoleh pelayanan kesehatan” dan Pasal 34 Ayat (3) yang berbunyi “Negara bertanggung jawab atas penyediaan fasilitas pelayanan kesehatan dan fasilitas pelayanan umum yang layak” [1].

Salah satu persoalan yang dihadapi BPJS Kesehatan yaitu banyaknya peserta yang menunggak dalam melakukan pembayaran sehingga menimbulkan defisit yang sangat besar pada BPJS Kesehatan dan menyebabkan banyaknya keluhan masyarakat terhadap BPJS Kesehatan. Menurut Kepala Biro Komunikasi dan Layanan Informasi (Kepala Biro KLI) Kementerian Keuangan (Kemenkeu) Nufransa Wira Sakti mengklarifikasi diantara penyebab utama defisit program Jaminan Kesehatan Nasional (JKN) yang sudah terjadi sejak awal pelaksanaannya adalah besaran iuran yang *underpriced* dan *adverse selection* pada peserta mandiri. Pada akhir tahun anggaran 2018, tingkat keaktifan peserta mandiri hanya 53,7%. Artinya, 46,3% dari peserta mandiri tidak disiplin

membayar iuran dengan kata lain menunggak. Sejak 2016 sampai dengan 2018, besar tunggakan peserta mandiri mencapai sekitar Rp. 15 triliun. Sepanjang 2018, total iuran dari peserta mandiri adalah Rp. 8,9 triliun, namun total klaimnya mencapai Rp. 27,9 triliun. Dengan kata lain, *claim ratio* dari peserta mandiri ini mencapai 313%. Dengan demikian, seharusnya kenaikan iuran peserta mandiri lebih dari 300% [2].

Penelitian sebelumnya yang dilakukan oleh Raya, Nurbaiti, & Sofia [3] tentang Klasifikasi Sentimen Masyarakat terhadap Kenaikan Harga Tiket Pesawat pada Twitter Menggunakan Naive Bayes. Hasil dari penelitian ini adalah kenaikan harga tiket pesawat dapat di klasifikasi dengan baik ke dalam bentuk positif dan negatif. 62% masyarakat menganggap kenaikan harga tiket pesawat sebagai hal yang biasa saja., 16% memberikan tanggapan positif dan 22% memberikan tanggapan negatif. Akurasi data latih Naive Bayes yaitu sebesar 100% dan *K-Nearest Neighbor Classifier* sebesar 73,33%. Untuk akurasi data uji Naive Bayes yaitu sebesar 90,70% dan *K-Nearest Neighbour Classifier* sebesar 62,79%. Dari hasil penelitian tersebut dapat disimpulkan bahwa algoritma Naive Bayes menghasilkan akurasi yang lebih tinggi dibandingkan KNN.

Text mining merupakan suatu proses pengetahuan intensif antara pengguna yang berinteraksi dan bekerja dengan sekumpulan dokumen atau data dengan menggunakan beberapa peralatan untuk menganalisis [4]. *Text mining* atau *text analytics* merupakan istilah yang biasa digunakan dalam mendeskripsikan sebuah teknologi yang mampu menganalisis data teks semi-terstruktur maupun tidak terstruktur. Adapun perbedaan *data mining* dengan *text mining* adalah *data mining* mengolah data yang bersifat terstruktur sedangkan *text mining* mengolah data atau dokumen semi-terstruktur maupun tidak terstruktur [5].

Algoritma *Naive Bayes Classifier* adalah algoritma yang mengklasifikasikan turunan dari teorema *Bayes* yang sederhana. Algoritma *Naive Bayes Classifier* digunakan untuk mencari nilai probabilitas tertinggi dalam mengklasifikasikan data *testing* pada kategori yang paling tepat. Algoritma *Naive Bayes Classifier* memiliki keunggulan yang sederhana, cepat, dan tingkat akurasi yang cukup tinggi daripada algoritma klasifikasi lainnya. Dalam penelitian ini yang menjadi data *testing* adalah dokumen test digital dari media sosial, yaitu tweet dari pengguna Twitter. *Naive Bayes Classifier* mengasumsikan bahwa keberadaan maupun ketidakberadaan sebuah fitur dalam sebuah kelas tidak memiliki keterkaitan dengan keberadaan maupun ketidakberadaan fitur lainnya [6]. Secara umum teorema Bayes dapat dinotasikan pada persamaan sebagai berikut:

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

Peluang kejadian A sebagai B ditentukan dari peluang B saat A, peluang A dan peluang B. $P(A|B)$ memiliki arti yaitu peluang A jika diketahui keadaan B.

Confusion matrix adalah suatu tabel untuk mengukur kinerja algoritma klasifikasi atau model klasifikasi atau *classifier*. Pada *confusion matrix* terdapat beberapa istilah yang umum digunakan, yaitu antara lain sebagai berikut:

1. *True Positives* (TP) merupakan persentase dari kelas

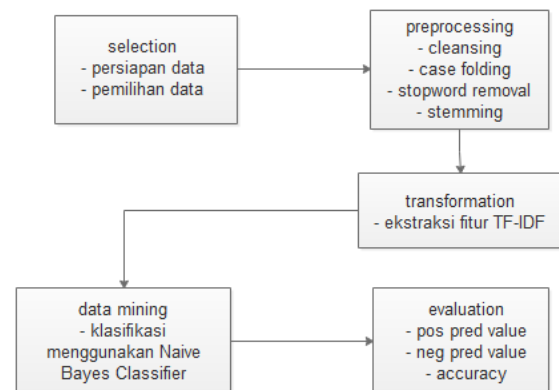
2. *True Negatives* (TN) merupakan persentase dari kelas negatif yang berhasil diklasifikasi sebagai kelas negatif.
3. *False Positives* (FP) merupakan kelas negatif yang diklasifikasi sebagai kelas positif.
4. *False Negatives* (FN) merupakan kelas positif yang diklasifikasi sebagai kelas negatif.

Term Frequency Inverse Document Frequency (TF-IDF) adalah sebuah metode yang digunakan untuk mengekstrak kalimat dengan cara memberikan nilai atau bobot pada kalimat. TF-IDF melakukan pembobotan terhadap kemunculan kata dalam suatu dokumen. TF menyatakan jumlah kata yang muncul dalam suatu dokumen. Sedangkan IDF menunjukkan tingkat kepentingan suatu kata yang terdapat dalam kumpulan dokumen. Pada penelitian sebelumnya nilai TF didapat dari nilai frekuensi kemunculan fitur t pada dokumen d .

$$TF_{td} = f(t, d)$$

II. METODOLOGI

Metodologi pada penelitian ini dengan mengikuti tahapan yang terdapat pada proses *data mining* yakni tahapan *Knowledge Discovery in Database* (KDD). Adapun alur atau tahapan dari KDD yakni sebagai berikut:



Gambar 1. Metodologi Penelitian

Pada Gambar 1 menjelaskan alur dalam pengolahan data yang akan di lakukan dalam penelitian ini yaitu:

1. *Selection*
2. *Preprocessing*
3. *Transformation*
4. *Data Mining*
5. *Evaluation*

III. HASIL DAN PEMBAHASAN

Data yang akan digunakan dalam proses klasifikasi sentimen terhadap kenaikan iuran BPJS Kesehatan pada *twitter*:

A. Selection

TABEL 1. JUMLAH TWEET SENTIMEN KENAikan IURAN BPJS KESEHATAN HASIL SELEKSI

Kata Kunci	Sept	Okt	Nov	Des	Jan	Feb
#bpjsnaik	3	2		3	8	
#iuranbpjs	2					
#BoikotBPJS			333	1	2	2
Total						356

Berdasarkan Tabel 1 dapat dijelaskan bahwa dari 5.821 data *tweet* didapatkan 356 *tweet* yang diperoleh dari bulan september hingga bulan februari. Penentuan kelas sentimen pada setiap data komentar dilakukan secara manual oleh peneliti, kemudian dilakukan tahap validasi oleh Ahli Bahasa Indonesia yaitu Bapak Imam Muhtarom, S.S., M.Hum yang merupakan Dosen Pendidikan Bahasa dan Sastra Indonesia Universitas Singaperbangsa Karawang. Berikut merupakan data sentimen yang telah di seleksi oleh peneliti dan telah dilakukan validasi oleh Ahli Bahasa Indonesia:

tweet	Label
Kalo naik sampe 2kali lipat di seluruh indonesia sement...	Negatif
Urus dulu @BPJSKesehatanRI boss, perbaiki manageme...	Negatif
rakyat dikorbankan #BoikotBPJS #BPJSMenyusahkanRak...	Negatif
Coba riset, survey dan buat laporan atas fakta dilapang...	Negatif
Boikot aja bpjs, biarin ga bisa bikin sim. Jangankan dian...	Negatif
Kt nya meringakan. Knp membebankan...? Kita di bilang "...	Negatif
#BPJSnaik Pelayanan harus naik atau ditingkatkan. #NKRI	Positif
Jangan katakan apa yang bisa diberikan negara kepada ...	Positif
yah dimaklumi tapi tetap	Positif
Gw sangat tertolong sih ada bpjs kesehatan ini..100rb p...	Positif
Harusnya bukan #BoikotBPJS, tapi #PerbaikiBPJS. Banya...	Positif
Bangke emng mereka yg menaikan iuran BPJS demi kepe...	Negatif
Rncana gue #BoikotBPJS Kalian digajih utk mensejahtra...	Negatif
biarin dah klo mau bangkrut juga #BoikotBPJS	Negatif
Pilih bpjs yg bangkrut atau rakyat yg bangkrut...? #Bub...	Negatif
Mau stop gak bisa, bubarkan saja BPJS. #BubarkanBPJS ...	Negatif

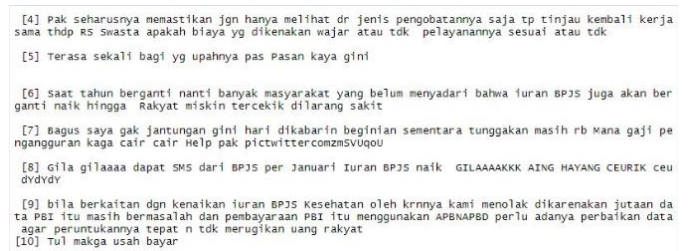
Gambar 2. Data Tweet Hasil Seleksi Ahli Bahasa

Berdasarkan Gambar 2 dapat dijelaskan bahwa data yang berhasil diseleksi oleh Ahli Bahasa Indonesia dibagi kedalam dua kelas yaitu kelas sentimen positif dan kelas sentimen negatif dengan jumlah data sebanyak 356 data *tweet* sentimen yang terdiri dari 26 data *tweet* untuk kelas sentimen positif dan 330 data *tweet* untuk kelas sentimen negatif.

B. Preprocessing

Data sentimen yang telah melalui tahap seleksi oleh Ahli Bahasa Indonesia, kemudian dilakukan tahap *preprocessing* untuk menghilangkan kata yang kurang berpengaruh dalam proses pengklasifikasian. Berikut merupakan hasil dari *preprocessing* data yang akan dilakukan.

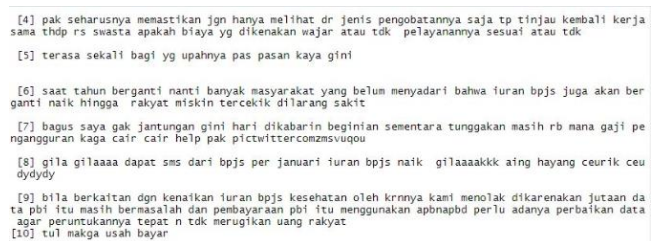
1. Cleansing



Gambar 3. Hasil Proses *Cleansing*

Dari Gambar 3 dapat dijelaskan bahwa data yang sudah dilakukan proses *cleansing* sudah berhasil menghilangkan *mention*, *hashtag*, dan tanda baca (*punctuation*) yang tidak berpengaruh.

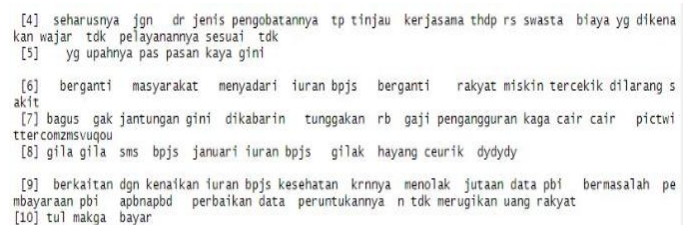
2. Case Folding



Gambar 4. Hasil Proses *Case Folding*

Pada Gambar 4 telah melalui tahap *cleansing* dimana semua data telah diubah menjadi huruf kecil.

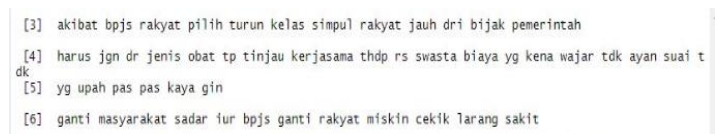
3. Stopword Removal



Gambar 5. Hasil Proses *Stopword Removal*

Dari Gambar 5 dapat dijelaskan bahwa data yang telah melalui tahap *stopword removal* telah dibandingkan dengan kosa kata yang terdapat dalam kamus *stopword*. *Stopword* yang digunakan dalam setiap studi kasus dapat disesuaikan sesuai kebutuhan data penelitian.

4. Stemming



Gambar 6. Hasil Proses *Stemming*

Dari Gambar 6 dapat dijelaskan bahwa data *tweet* yang telah melalui tahap *stemming* merupakan data bersih yang sudah dihilangkan imbuhan dan menjadi kata dasar sehingga dapat dilanjutkan ke tahap selanjutnya.

C. Transformation

	ayan	harus	tingkat	bayar	jangan	rakyat	akibat	bijak	bpjs
1	1.7179351	1.2059175	1.96359	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
2	0.0000000	0.0000000	0.000000	0.86436346	1.3611387	0.37862781	0.0000000	0.0000000	0.0000000
3	0.0000000	0.0000000	0.000000	0.00000000	0.0000000	0.18931390	0.7063111	0.4180252	0.07360637
4	0.3435870	0.2411835	0.000000	0.00000000	0.0000000	0.00000000	0.0000000	0.0000000	0.00000000
5	0.0000000	0.0000000	0.000000	0.00000000	0.0000000	0.00000000	0.0000000	0.0000000	0.00000000
8	0.0000000	0.0000000	0.000000	0.00000000	0.0000000	0.00000000	0.0000000	0.0000000	0.16059571
10	0.0000000	0.0000000	0.000000	0.86436346	0.0000000	0.00000000	0.0000000	0.0000000	0.00000000
11	0.5153805	0.0000000	0.000000	0.00000000	0.0000000	0.00000000	0.0000000	0.0000000	0.00000000
12	0.0000000	0.0000000	0.000000	0.00000000	0.8166832	0.22717669	0.0000000	0.0000000	0.00000000
13	0.0000000	0.0000000	0.000000	0.00000000	0.0000000	0.00000000	0.0000000	0.0000000	0.00000000

Gambar 7. Bobot Pada Data Training

Proses pembobotan dilakukan dengan membandingkan bobot setiap kata pada *data testing* dengan kata pada *data training*. Jika bobot probabilitas negatif lebih besar, maka hasil *tweet* tersebut kelas negatif. Namun, jika bobot probabilitas positif lebih besar, maka hasil *tweet* tersebut kelas positif.

D. Data Mining

Pada tahap *data mining* adalah tahap dilakukannya klasifikasi data teks menggunakan *Naive Bayes Classifier*. Pada penelitian ini dilakukan lima kali pengujian yaitu menggunakan *persentase* 90% data latih dan 10% data uji. Pengujian yang kedua menggunakan *persentase* 80% data latih dan 20% data uji. Pengujian ketiga menggunakan *persentase* 70% data latih dan 30% data uji. Pengujian keempat menggunakan *persentase* 60% data latih dan 40% data uji. Pengujian kelima menggunakan *persentase* 50% data latih dan 50% data uji [7].

TABEL 2. PEMBAGIAN DATA LATIH DAN DATA UJI

Split Ratio	Persentase (%)	Data Latih	Data Uji
0,9	90% dan 10%	320	36
0,8	80% dan 20%	285	71
0,7	70% dan 30%	249	107
0,6	60% dan 40%	214	142
0,5	50% dan 50%	178	178

Berikut merupakan proses *training* dengan mengklasifikasikan data latih secara manual yang telah melalui proses *preprocessing*, yaitu sebagai berikut:

TABEL 3 DATA TRAINING

Tweet	Fitur	Class
Tweet 1	bpjs pemerasan sistem kompeni	Negatif
Tweet 2	boikot bpjs lintah darat ngisepin darah rakyat	Negatif
Tweet 3	pemaksaan bayar iur bpjs radikal sampe pake tukang tagih	Negatif
Tweet 4	gotong royong tolong sehat bantu sakit moga lambung sadar	Positif
Tweet 5	naik iur bpjs sehat jangan pandang negatif banding lurus tingkat kualitas sehat moga	Positif

Tweet 6	pasien sakit jantung rugi kalo bayar uang obat untung bpjs	Positif
---------	--	---------

Berdasarkan Tabel 3, maka dapat dibuat sebuah model *probabilitas* sebagai berikut:

$$P(V_j) = \frac{docs_j}{training} = \frac{3}{6} = 0,5 \tag{1}$$

$$P(a_{bpjs}|V_{negatif}) = \frac{3+1}{18+42} = 0,067 \tag{2}$$

$$P(a_{bpjs}|V_{positif}) = \frac{2+1}{28+42} = 0,043 \tag{3}$$

$$P(a_{pemerasan}|V_{negatif}) = \frac{1+1}{18+42} = 0,033 \tag{4}$$

$$P(a_{pemerasan}|V_{positif}) = \frac{0+1}{28+42} = 0,014 \tag{5}$$

$$P(a_{sistem}|V_{negatif}) = \frac{1+1}{18+42} = 0,033 \tag{6}$$

$$P(a_{sistem}|V_{positif}) = \frac{0+1}{28+42} = 0,014 \tag{7}$$

$$P(a_{kompeni}|V_{negatif}) = \frac{1+1}{18+42} = 0,033 \tag{8}$$

$$P(a_{kompeni}|V_{positif}) = \frac{0+1}{28+42} = 0,014 \tag{9}$$

E. Evaluation

Berdasarkan percobaan diatas dapat dijelaskan bahwa hasil akurasi menggunakan algoritma *Naive Bayes Classifier* dengan seleksi fitur TF-IDF maka didapatkan model dengan nilai *accuracy* 92,70%.

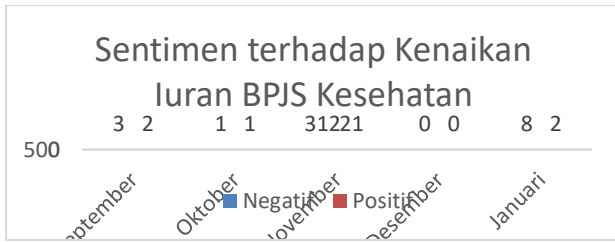
Berikut merupakan hasil akurasi terhadap klasifikasi sentimen kenaikan iuran BPJS Kesehatan, yaitu sebagai berikut:

TABEL 4. TABEL AKURASI MODEL KLASIFIKASI

Rasio Split	Accuracy (%)	95% CI	Rentang CI
50:50	92,70%	0,8783:0,9605	0,08220
60:40	92,96%	0,8743:0,9657	0,09140
70:30	92,52%	0,8580:0,9672	0,01092
80:20	92,96%	0,8433:0,9767	0,01334
90:10	91,67%	0,7753:0,9825	0,02072

Berdasarkan Tabel 4 dapat dijelaskan bahwa terdapat lima skenario pembagian data *training* dan data *testing* dengan hasil yang telah ditunjukkan pada Tabel 4 yaitu pada rasio *split* 60:40 merupakan hasil dengan akurasi tertinggi sebesar 92,96% dengan rentang kelebaran CI (*Confidence Interval*) lebih sempit dibandingkan dengan rasio *split* 80:20 [8].

Berdasarkan hasil seleksi yang dilakukan oleh Ahli Bahasa Indonesia, data sentimen terhadap kenaikan iuran BPJS Kesehatan berjumlah 356 *tweet* yang terdiri dari 330 *tweet* negatif dan 26 *tweet* positif.

Gambar 8. Bobot Pada *Data Training*

Berdasarkan grafik pada Gambar 8 dapat dijelaskan bahwa data *tweet* sentimen terhadap kenaikan iuran BPJS Kesehatan lebih dominan terhadap kelas sentimen negatif dengan total data sentimen negatif yaitu sebesar 330 *tweet* yang tersebar pada bulan September, Oktober, November, dan Januari dimana *tweet* terbanyak ada pada bulan November yaitu sebesar 312 *tweet* sentimen negatif. Sedangkan *tweet* sentimen positif hanya terdapat 26 data *tweet* dengan populasi terbanyak ada pada bulan November yaitu 21 *tweet*.

Klasifikasi data sentimen terhadap kenaikan iuran BPJS Kesehatan pada *twitter* menggunakan *Naive Bayes Classifier* dilakukan dengan lima pengujian dan mendapatkan akurasi tertinggi pada rasio *split* 60:40 sebesar 92,96% dengan pembagian data yakni 71 data uji dan rentang kelebaran 95% CI (*Confidence Interval*) sebesar 0,0914 yang menghasilkan presisi cukup baik [8] karena memiliki rentang lebih sempit dibandingkan dengan pengujian pada rasio *split* 80:20.

Secara keseluruhan dapat disimpulkan bahwa algoritma *Naive Bayes Classifier* pada penelitian ini menghasilkan akurasi yang tinggi.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan terdapat kesimpulan yaitu Analisis sentimen terhadap kenaikan iuran BPJS Kesehatan dilakukan dengan membagi data yang telah melewati tahap *preprocessing* menjadi data latih dan data uji.

Pengujian dilakukan sebanyak 5 kali dengan *persentase* 90% data latih dan 10% data uji, 80% data latih dan 20% data uji, 70% data latih dan 30% data uji, 60% data latih dan 40% data uji, dan 50% data latih dan 50% data uji. Hasil klasifikasi 320 data latih dan 36 data uji menggunakan *Naive Bayes Classifier* dengan seleksi fitur TF-IDF pada Rstudio didapatkan 33 data kelas sentimen negatif diprediksi benar sebagai kelas sentimen negatif, sedangkan 3 data kelas sentimen positif diprediksi sebagai kelas sentimen negatif. Hasil akurasi terbaik merupakan klasifikasi menggunakan *Naive Bayes Classifier* dengan seleksi fitur TF-IDF yaitu sebesar 92,96% dengan *persentase* data sebanyak 60% data latih dan 40% data uji.

ACKNOWLEDGMENT

Alhamdulillah, segala puji serta syukur ke hadirat Allah SWT atas karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan penelitian ini. Terima kasih kepada Universitas Singaperbangsa Karawang atas izin dan dukungan dalam proses penelitian ini. Serta terima kasih atas semua do'a, kasih sayang, nasihat, semangat serta

dukungan dari Orang tua, keluarga dan sahabat yang telah diberikan sampai saat ini.

REFERENCES

- [1] D. Septia, "Kepuasan Terhadap Kualitas Pelayanan BPJS Kesehatan," 19 Desember 2017. [Online]. Available: <http://digilib.unila.ac.id>.
- [2] K. K. R. Indonesia, "Mengapa Iuran BPJS Naik ?," 09S September 2019. [Online]. Available: <https://www.kemenkeu.go.id>.
- [3] A. M. Raya, F. Nurbaiti and D. Sofia, "Klasifikasi Sentimen Masyarakat terhadap Kenaikan Harga Tiket Pesawat pada Twitter Menggunakan Naive Bayes," *Jurnal Instek*, pp. 239-246, 2019.
- [4] R. Feldman and J. Sanger, *The Text Mining Handbook Advanced Approaches In Analyzing Unstructured ed Data*, New York: Cambridge University Press, 2007.
- [5] H. N. Jamil, "Analisis Sentimen Pada Online Review Menggunakan Kombinasi Metode Lexicon Based dan Naive Bayes Classifier," Skripsi. Program Studi Statistika FMIPA UII Yogyakarta, Yogyakarta, 2017.
- [6] A. Syakuro, "Analisis Sentimen Masyarakat terhadap E-Commerce pada Media Sosial Menggunakan Metode Naive Bayes Classifier (NBC) dengan Seleksi Fitur Information Gain (IG)," Universitas Islam Negeri Maulana Malik Ibrahim, Malang, 2017.
- [7] M. S. Simatupang, "Analisis Sentimen Pada Komentar Body Shaming Beauty Vlog Youtube," Karawang, 2019.
- [8] N. Ms, "Estimasi-Confidence Interval," 2019. [Online]. Available: academia.edu.