Analisis Sentimen Publik Terhadap Calon Presiden 2019 Melalui Twitter Menggunakan Metode Naive Bayes Classifier (Studi kasus : Pilpres 2019)

Bonar Panjaitan[1], Kemas Muslim Lhaksmana[2]

[1,2]Fakultas Informatika, Universitas Telkom, Bandung
[1]bonarpanjaitan@students.telkomuniversity.ac.id, [2]kemasmuslim@telkomuniversity.ac.id

Abstrak

Pemilihan umum (Pemilu) merupakan bagian dari demokrasi suatu negara termasuk Indonesia. Pemilihan dilakukan oleh rakyat yang telah memenuhi syarat sebagai pemilih. Mengingat pengaruh media sosial yang cukup pesat dikalangan masyarakat, maka dari itu penulis tertarik untuk mengetahui opini publik atau pemikiran rakyat yang dicurahkan kedalam akun media sosial mereka. Dengan mengetahui sentimen positif atau negatif, penulis berharap akan mendapatkan akurasi yang tinggi pada suatu pembicaraan berupa *tweet* atau komentar masyarakat yang ditujukan pada pasangan calon presiden 2019. Tujuan dari penelitian ini sekaligus menguji akurasi *Naive Bayes Classifier* pada proses perhitungan probabilitas pada kasus yang telah dipilih. Untuk mendukung perhitungan naive bayes, peneliti akan melakukan tahap *preprocessing* telebih dahulu sebelum menggunakan TF-IDF(*Term Frequency-Inverse Document Frequency*) sebagai metode pembobotan agar memberikan perhitungan akurasi yang lebih baik [1]. Dalam penelitian ini, penulis mendapatkan hasil akurasi 57,00%. Hasil ini sudah cukup baik dalam klasifikasi positif dan negatif sebuah dokumen atau kalimat.

Kata kunci : ***Preprocessing, Twitter, Naive Bayes Classifier.***

Abstract

General election are part of a country's democracy, including Indonesia. Elections are conducted by people who have fulfilled the requirements as a voters. Considering the influence of social media is quite rapid among the people, therefore the author are interested of knowing public opinion or people thoughts that poured out into their social media account. By knowing positive or negative sentiments, the author hopes to get high accuracy in a conversation in the form of tweets as well as public comments aimed to 2019 presidential candidates. The purpose of this study while testing accuracy of Naive Bayes Classifier in the process of calculating probability of the selected case. To support Naive Bayes Classifier calculation, the author will do the preprocessing as the first step before using TF-IDF(Term Frequency-Inverse Document Frequency) as a weighting method to provide a better calculation of accuracy [1]. In this study, the authors obtained an accuracy of 57.00%. This result is good enough in the positive and negative classification of a document or sentence.

Keywords: ***Preprocessing, Twitter, Naive Bayes Classifier.***

1. Pendahuluan

Seiring dengan perkembangan teknologi informasi menjadikan internet sebagai hal yang paling dibutuhkan oleh masyarakat. Sehingga muncul media baru untuk menyebarkan informasi melalui internet kepada khalayak yaitu media online. Salah satu media online yang digunakan hingga saat ini adalah media sosial.

Media sosial digunakan sebagai alat komunikasi secara tidak langsung. Selain sebagai alat komunikasi, media sosial digunakan untuk mendapatkan informasi, hiburan, pendidikan, dan akses pengetahuan dari berbagai tempat yang berbeda. Karena media sosial memiliki berbagai fungsi yang bermanfaat, maka banyak masyarakat *cyber* menggunakan media sosial, khususnya di Indonesia.

Sesuai dengan data hasil survey dari www.apjii.or.id jumah penduduk Indonesia yang menggunakan media sosial mencapai 129,2 juta atau 97,4% dari total penduduk Indonesia salah satunya adalah Twitter. Twitter bermanfaat untuk menyampaikan tawaran atau memberitakan peristiwa, mempromosikan konten terbaru dan menghubungkan para follower dengan tautan-tautan berisi berita penting. Di twitter pengguna bisa menjalin jaringan dengan pengguna lain, menyebarkan informasi, mempromosikan pendapat pengguna lain, membahas isu terhangat (*trending topic*) dan menjadi bagian dari isu tersebut dengan menggunakan *hashtag* tertentu. Dilansir dari www.cnnindonesia.com, bahwa CEO Twitter Dick Costolo akhirnya

mengungkap jumlah pengguna twitter di Indonesia yang jumlahnya mencapai 50 juta pengguna dan ia yakin angka itu akan terus bertambah di masa depan.

Ia mengklaim twitter juga memberikan banyak keuntungan kepada konsumen di Indonesia karena menghubungkan banyak orang sampai menjadi wadah untuk membicarakan hal yang sedang terjadi salah satunya adalah membahas tentang politik. Penulis tertarik tentang bagaimana sentimen publik terhadap calon presiden. Data-data dari *Twitter* akan diolah menggunakan teknik *data mining*. Sentimen ini akan diklasifikasikan menjadi tiga bagian, yaitu sentimen negatif, , dan positif berdasarkan dari *tweet* yang dilontarkan oleh para pengguna media sosial yang berhubungan dengan calon presiden. Klasifikasi akan dilakukan menggunakan metode Naïve Bayes *Classifier* [2][9].

2. Studi Terkait

Pada penelitian tugas akhir ini, analisis sentimen menggunakan metode *Naive Bayes Classifier* akan menggunakan sejumlah 500 data guna untuk menghasilkan akurasi yang diharapkan. Klasifikasi sentimen akan menggunakan data latih yang telah di *crawl* dari *twitter* [4]. Penelitian yang berhubungan dengan tugas akhir ini sebelumnya adalah milik jaka sembodo pada tahun 2016 tentang "Data Crawling Otomatis pada Twitter" [11].

2.1 Analisis sentimen

Sentiment Analisys merupakan salah satu bagian dari Natural Language Processing (NLP) untuk mengetahui kondisi dari populasi dunia maupun dalam ruang lingkup lebih kecil, mengenai sebuah topik bahasan yang sedang hangat dibicarakan. Sentimen analisis juga dengan *opinion mining* yang berarti mengumpulkan dan mengolah data menjadi suatu infomasi. Sentimen analisis dapat dilakukan dengan media artikel pada blog, komentar seseorang pada sebuah forum, review ataupun *tweet*. Analisis sentiment sendiri merupakan proses dalam mengekstrak data sentimen yang akan dikategorikan menjadi tiga bagian, yaitu: positif, negatif. Berikut contoh kalimatnya adalah, "Kamu akan tidak akan mampu"(negatif), dan "Tahan, kita akan menang"(positif) [8][10].

2.2 *Preprocesing*

Pre-processing data merupakan langkah dalam menyiapkan data yang sesuai dengan kebutuhan secara efisien sehingga dapat diproses dengan menggunakan metode pada bagian utama sistem. Tahapan dalam melakukan pre-*processing* data adalah sebagai berikut: case folding, tokenizing, filtering, dan stemming [3].

2.3 TF-IDF

Tf-Idf merupakan metode dalam pembobotan kata yang berarti Term Frequency-Inverse Document Frequency yang biasanya digunakan dalam mencari informasi. TF-IDF merupakan ukuran statistik yang digunakan untuk mengvaluasi seberapa penting sebuah kata pada sebuah dokumen atau dalam sebuah kalimat. *Term Frequency* digunakan untuk menghitung berapa banyak sebuah kata muncul pada sebuah dokumen sedangkan *Inverse Document Frequency* digunakan sebagai perbandingan dari kata yang didapat dari *Term Frequency* [13].

$$IDF = log \frac{D}{Df_i} \quad (2.1)$$

Keterangan:
IDF = Nilai *inverse* dari DFi
D = Banyaknya *tweet* pada datasets
Dfi = Banyaknya *tweet* pada datasets yang mengandung kata ke-i

Maka nilai TF-IDF adalah:

$$TF - IDF = TF * IDF \quad (2.2)$$

Keterangan:
TF-IDF = Nilai bobot kata dalam sebuah datasets
TF = Nilai jumlah kemunculan sebuah kata pada *tweet*
IDF = Nila kata yang muncul dalam sebuah datasets

2.4 NBC

Teorema Bayes merupakan sebuah pendekatan pada ketidak-tentuan yang diukur dengan probabilitas. Teorema Bayes ditemukan pada abad ke-18 oleh seorang ahli probabilitas yaitu Thomas Bayes. Teorema ini digunakan untuk mengklasifikasikan kelas-kelas dengan menggunakan probabilitas atau lebih dikenal dengan nama *Naïve Bayes Classifier*(NBC). Metode NBC menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan akan dilakukan proses analisis terhadap sampel dokumen berupa vocabulary. Selanjutnya penentuan probabilitas prior untuk tiap kategori dari suatu dokumen berdasarkan term yang muncul dalam dokumen yang diklasifikasi.

Mengasumsikan bahwa algoritma *atribute* objek yang digunakan bersifat independen, sedangkan probabilitas yang menggunakan perkiraan akhir dihitung sebagai jumlah frekuensi dari *table* keputusan [12][15]. Untuk menghitung probabilitas sentimen menggunakan rumus dibawah ini :

$$V_{MAP} = \underset{V_j eV}{\arg max} \prod_{i=1}^{n} P(x_i|V_j)P(V_j) \quad (2.3)$$

Sedangkan untuk penggunakan *Laplace smooting* pada naive bayes menggunakan rumus di bawah ini :

$$P(t_k|c) = \frac{w_{ct}+1}{(\Sigma_{w'\epsilon v}W'_{ct})+B'} \quad (2.4)$$

w_{ct} = Pembobotan TF-IDF.
$\Sigma_{w'\epsilon v}W'_{ct}$ = Jumlah total W dari keseluruhan kata yang berada di kelas c.
B' = Jumlah total W kata unik di semua kelas.

2.5 Measuring Performance

Measuring performance merupakan salah satu metode dari statistika yang digunakan untuk mengevaluasi juga membandingkan algoritma learning dengan membagi data menjadi 2 segmentasi, data pertama digunakan untuk training sebuah model dan data kedua akan digunakan untuk memvalidasi model tersebut. Selanjutnya adalah cross validation yang bertujuan untuk mendapatkan performansi dari model tersebut. Dalam mempermudah perhitungan nilai akurasi peneliti akan menggunakan *confusion matrix.* Berikut bentuk umum dari *confusion matrix.*

Tabel 1. Nilai *Confusion Matrix*

Actual/Classified	*Classified Positif*	*Classified Negatif*
Actual Positif	*True Positif*(TP)	*False Negatif*(FN)
Actual Negatif	*False Positif*(FP)	*True Negatif*(TN)

Berikut yang dapat diukur oleh *Confusion Matrix* :

a. Recall
Recall adalah perbandingan hasil klasifikasi dengan kelas sesungguhnya atau tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

$$Recall = \frac{\text{TP}}{\text{TP+FN}} \quad (2.5)$$

b. Precision
Precision adalah perbandingan antara data yang terdeteksi benar dengan seluruh data prediksi pada suatu kelas atau tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \quad (2.6)$$

c. Accuracy
Accuracy adalah perbandingan antara data yang terdeteksi benar dengan seluruh data hasil prediksi.

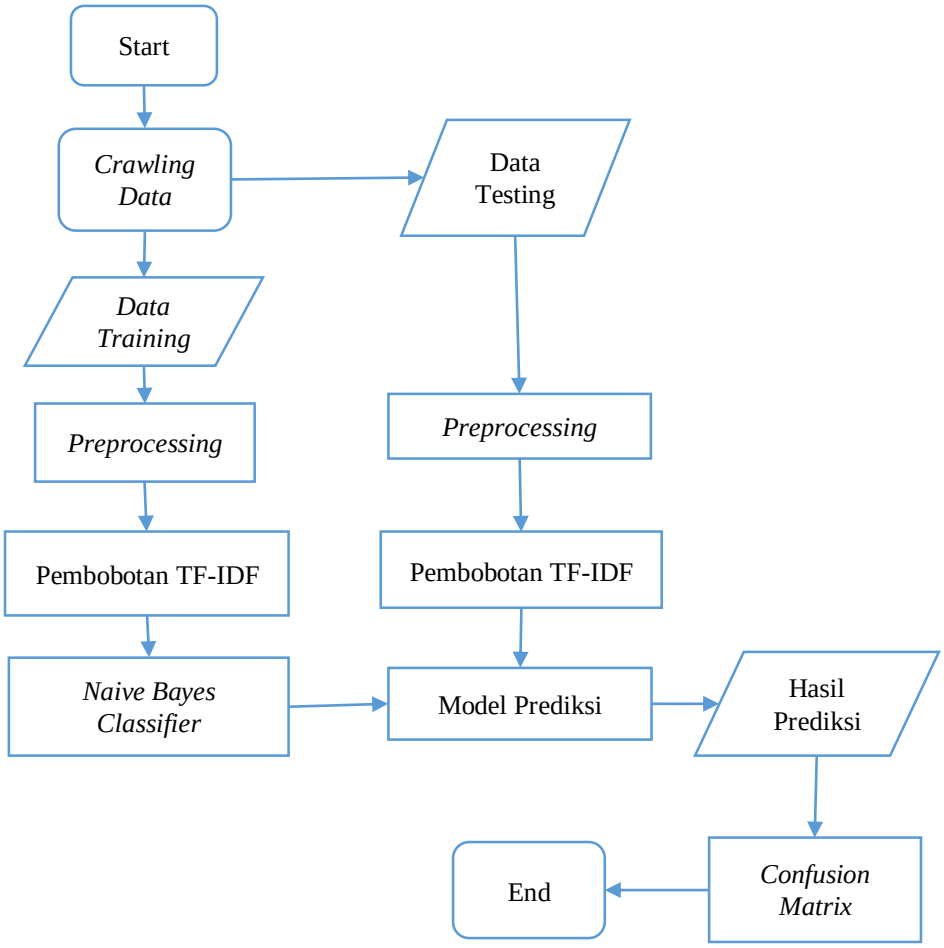
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad 2.5$$

3. Sistem yang Dibangun

Pada perancangan sistem, peneliti akan membahas tentang tujuan sistem yang akan digunakan untuk memprediksi kecenderungan politik seseorang berdasarkan *tweet* yang dilontarkan pengguna. Tahap pertama yang akan dilakukan adalah pre-processing data yang bertujuan untuk menyederhanakan data sehingga lebih mudah pada saat pengolahan data dengan menggunakan metode NBC.

3.1 Rancangan Sistem

Penelitian ini bertujuan untuk mengklasifikasikan sentimen para pengguna yang terdapat pada twitter dan berhubungan dengan pemilihan umum 2019. Klasifikasi akan dibagi menjadi beberapa jenis kategori yaitu sentimen positif dan negatif. Metode yang akan digunakan pada klasifikasi sentimen adalah metode Naïve Bayes *Classifier*. Berikut rangkaian proses pada sistem yang akan dibangun.



Gambar 1. Flowchart Sistem

a. Pengumpulan Data

Pengumpulan data dengan melalui API yang disedikan oleh *Twitter*, diambil dengan cara *crawling* menggunakan *data crawler*, Mendapatkan sekitar 500 data yang berasal dari data *Tweet* dan *comment* di *Twitter* [11].

Tabel 2. Data *Twitter* hasil *crawling*.

No Urut	*Comment*	*Tweet*
1	Semangat pak, semoga bisa 2 periode seluruh rakyat Indonesia percaya sama bapak karna kesederhanaan dan semua yang sudah bapak realisasikan	karding sangat menyayangkan pernyataan prabowo subianto yang menolak memercayai hasil pemilu 2019. penolakan itu tidak dibangun atas data dan fakta
2	Alhamdulillah, Jokowi-Amin bisa menjadikan Indonesia lebih maju dan lebih bermartabat	@MayjenSudrajat Insya Allah pak Jokowi mudah mudahan Indonesia lebih maju

3	Saat itu Prabowo menjabat sebagai pemimpin Komando Pasukan Khusus atau Kopassus. Ia dan pasukannya sedang berada di daerah rawan, sehingga helikopter tak berani mendarat.	RT @MahesaTiwi: @Deddy_Mizwar_ mendukung @jokowi untuk menang 2 periode! Selain itu banyak kader @PDemokrat pula yang merapatkan barisan ke jokowi

b. *Preprocessing*

Dalam membuat keputusan yang baik, maka harus menggunakan data yang baik pula (lengkap, benar, konsisten, dan terintegrasi). Sebelum melakukan data mining, perlu dilakukan pre-*processing* data terlebih dahulu untuk memastikan data yang akan diolah di data mining adalah data yang baik. Data yang kualitasnya kurang baik, dapat disebabkan oleh beberapa hal, yaitu: Data tidak lengkap, *noisy* (ada data yang berbeda sendiri), tidak konsisten (tidak sesuai dengan rule yang ditentukan). Untuk mengatasi masalah tersebut, maka dilakukanlah *pre-processing* data sebelum diolah dengan data mining [3]. Pre-*processing* dapat dilakukan dengan beberapa teknik, yaitu:

- *Cleaning* adalah memperkecil jumlah data yang hilang atau berbeda.
- Integrasi adalah menggabungkan beberapa sumber data sehingga dapat saling melengkapi.
- Transformasi adalah mengubah data kompleks dengan tidak menghilangkan isi, sehingga lebih mudah diolah.
- Diskretisasi adalah membagi nilai data menjadi beberapa range data

Cleaning adalah mengurangi jumlah data sehingga sumber yang digunakan lebih sedikit supaya prosesnya lebih cepat dilakukan.

Tabel 3. Hasil *preprocessing Twitter*

No Urut	*Sebelum Preprocessing*	*Setelah Preprocessing*
1	Mari bersama tolak people power #muijabar #tolakpeoplepower #2019tetapjokowi #jokowi #pilpres2019 #nkri #jokowimaruf #jokowi2periode #ingatjokowinawacita #debatpilpres2019 #kubuhoax #kecebong #kampret	mari bersama tolak people power muijabar tolak people power 2019 tetap jokowi jokowi pilpres 2019 nkri jokowi maruf jokowi 2 periode ingat jokowi nawacita debat pilpres2019 kubu hoax kecebong kampret
2	Sekarang siapa yang antek asing #2019pilihjokowi #2019jokowipresidenRI #jokowi #jambi	sekarang siapa yang antek asing 2019 pilih jokowi 2019 jokowi presiden ri jokowi jambi
3	kubu 01 aneh sekali bukan????	kubu 01 aneh sekali bukan
4	Kami tidak terima kalau kalah dalam keadaan dicurangi #prabowo #indomesiamaju	kami tidak terima kalau kalah dalam keadaan dicurangi prabowo indonesia maju

Selanjutnya adalah melakukan *stemming* pada setiap kata atau mengubah kata berimbuhan menjadi kata dasar. Tahap ini bertujuan untuk memaksimalkan perhitungan dalam prosesi metode *Naive Bayes Classifier*.

Tabel 4. Hasil *preprocessing Twitter*

No Urut	*Sebelum Preprocessing*	*Setelah Preprocessing*
1	ngomong gitu boleh perempuan berita sandiaga uno prabowo	omong gitu boleh perempuan dan berita sandiaga uno prabowo
2	waspada hati-hati jebakan betmen prabowo prabowo sandi	waspada hati hati jebak betmen prabowo prabowo sandi

c. Tokenizing

Tokenizing atau Tokenisasi, pada tahap ini kalimat akan dipecah menjadi kata berurutan. Kata akan dirubah menjadi 3 bentuk kata. Kata tersebut merupakan Unigram setiap kalimat dibagi menjadi satu kata. Bigram setiap kalimat dibagi menjadi dua kata. Trigram setiap kalimat dibagi menjadi tiga kata.

Tabel 5. Contoh Hasil Tokenisasi

No	Bentuk	Kata
1	Unigram	saya yakin indonesia akan menang saya optimis
2	Bigram	saya yakin, yakin indonesia, indonesia akan, akan menang, menang saya, saya optimis
3	Trigram	saya yakin indonesia, indonesia akan menang, menang saya optimis

d. Pembobotan

Pembobotan merupakan perhitungan frekuensi banyaknya jumlah kemunculan suatu kata atau term (TF tinggi) dalam dokumen, semakin besar pula bobotnya atau akan memberikan nilai kesesuaian yang semakin besar [5].

4. Hasil dan Pengujian

Pada tahap ini akan dijelaskan mengenai hasil dan pengujian yang dilakukan serta didapatkan.

4.1 Dataset dan *labelling*

Selanjutnya adalah dataset dan *labelling* menggunakan data *crawler* yang dilakukan dalam waktu kurang lebih satu tahun sejak 2018. Berikut kata kunci yang digunakan dalam *crawler*.

Tabel 6. Pencarian pada *Crawling*

Kata Kunci		
Pilpres2019	Cebong	Prabohong
Jokowi	Kampret	Debatcapres2019
Prabowo	Maruf Amin	Jokowow
Jokowi Maruf	Sandiaga Uno	Prabowopresidenku
Prabowo Sandi	Wiwiwowo	Jokowi2periode

Banyaknya jumlah dataset yang digunakan adalah 500 jumlah *tweet* dan data *comment*. Kemudian data *Tweet* dan *comment* tersebut dilabelkan secara manual. Tabel berikut merupakan kalimat yang akan dilabelkan yang berasal dari data *Twitter*.

Dari data yang telah didapatkan, akan dilakukan pelabelan data secara manual menggunakan syarat positif dan negatif yang telah ditentukan. Syarat tersebut akan dijadikan acuan untuk menentukan dokumen positif atau negatif. Ada 50 jumlah data syarat positif seperti : Bravo, alhamdulillah, semangat, jokowi, prabowo, dll. Sedangkan 50 jumlah data syarat negatif adalah : wiwi, wowo, prabohong, kampret, cebong, dll.

.

Tabel 7. *Labelling* data pada *Twitter*

No Urut	*Tweet*	Label
1	saya yakin joko widodo yang akan menang	Positif
2	prabohong tolak hasil pilpres 2019 mahfud md bisa jadi preseden buruk andai menang beneran di 2024 hasilpilpres2019	Negatif

Dari seluruh 500 data yang telah di labelkan manual, mendapatkan 361 data Positif, 139 data Negatif yang ditunjukan pada tabel 9.

Tabel 8. Jumlah hasil dari labeling setiap *tweet* pada seluruh data

Label	Jumlah Data
Positif	331

Negatif	169

4.2 Skenario Pengujian

Skenario pengujian bertujuan untuk mengetahui akurasi yang didapatkan pada percobaan. Pembobotan yang digunakan pada tahap ini adalah TF-IDF dan tanpa TF-IDF menggunakan data yang telah didapatkan dari *twitter***.**

Tabel 9.Skenario Pengujian

Parameter	Kode	Skenario
Akurasi Pengujian TF-IDF	JSPM 1	Pengujian bentuk kata Unigram dengan TF-IDF
	JSPM 2	Pengujian bentuk kata Bigram dengan TF-IDF
	JSPM 3	Pengujian bentuk kata Trigram dengan TF-IDF
	JSPM 4	Pengujian bentuk kata Unigram + Bigram dengan TF-IDF
	JSPM 5	Pengujian bentuk kata Unigram + Trigram dengan TF-IDF
	JSPM 6	Pengujian bentuk kata Bigram + Trigram dengan TF-IDF
	JSPM 7	Pengujian bentuk kata Unigram + Bigram + Trigram dengan TF-IDF

4.3 Hasil Pengujian

Pengujian pada skenario yang dibuat terlebih dahulu melakukan pengujian berupa pembagian data *training* dan data *testing* menggunakan seluruh data yang digunakan. Pengujian dilakukan menggunakan *confusion matrix* untuk mencari nilai akurasi pada percobaan.

Tabel 10. Akurasi Pengujian Seluruh Data

Perbandingan Data	Akurasi Total
90:10	57.00%
80:20	56.80%
70:30	56.60%
60:40	52.71%
50:50	49.34%

Pengujian 500 data *Twitter* dengan perbandingan 450 data latih dan 50 data uji yang di gunakan untuk mencari akurasi terbaik untuk semua skenario. 450 data latih dan 50 data uji tersebut dipilih secara acak atau randomisasi pemilihan data.

Tabel 11. Akurasi Hasil Pengujian data yang didapat melalui *Twitter*

Kode Skenario	Akurasi
JSPM 1	0.5560
JSPM 2	0.5680
JSPM 3	0.5660
JSPM 4	0.5660
JSPM 5	0.5560
JSPM 6	0.5700
JSPM 7	0.5700

Dari hasil pengujian semua skenario data *Twitter* tersebut mendapatkan akurasi terbaik data *Twitter* skenario pengujian kata dengan TF-IDF Unigram + Bigram + Trigram Jokowi Amin (JSPM 7) sebesar 57,00%.

Tabel 12. Akurasi, *Recall* dan *Precision* terbaik pada semua percobaan.

Kode Skenario	Akurasi	*Recall* Positif	*Recall* Negatif	*Precision* Positif	*Precision* Negatif
JSPM 7	0.5700	0.7523	0.2130	0.6518	0.3051

Dari pengujian Twitter setiap skenario didapatkan confusion matrix untuk skenario terbaik yaitu skenario pengujian TF-IDF Unigram + Bigram + Trigram Jokowi Amin (JSPM 7) dengan jumlah confusion matrix 249 *True Positif* , 82 *False Negatif*, 133 *False Positive*, 36 *True Negative*.

Tabel 13. *Confusion Matrix* dari skenario terbaik menggunakan data *Twitter*

Kelas JSPM	Kelas Prediksi	
	Positif	Negatif
Positif	249	82
Negatif	133	36

4.4 Analisis

Dari hasil pengujian perbandingan penggunaan jumlah data, dapat disimpulkan bahwa nilai akurasi akan semakin menurun seiring dengan berkurangnya jumlah data training. Hal ini terjadi karena semakin sedikit jumlah kata pada proses TF-IDF, maka akan semakin kecil akurasi yang diambil menggunakan data *training*.

Hasil pengujian diatas menunjukkan bahwa nilai TF-IDF mempengaruhi akurasi sistem klasifikasi yang dibuat oleh peneliti. Hal itu dapat dianalisis dari skenario JSPM 7 yang memiliki tingkat akurasi 57,00%. Dari hasil pengujian ini juga dapat dilihat bahwa gabungan bentuk kata dapat meningkatkan akurasi.

Pembatasan jumlah pada proses TF-IDF juga dapat mempengaruhi tingkat akurasi yang akan didapatkan. Jika jumlah kata pada TF-IDF berjumlah sedikit, maka hasil klasifikasi akan semakin tidak akurat karena beberapa kata saja yang dapat diproses dan dijadikan keunikan dari *labelling* tersebut.

Dari tabel *confusion matrix* juga dapat dilihat bahwa kalimat yang telah diprediksi dengan baik adalah *tweet* positif. Hal ini dikarenakan banyaknya kombinasi kata pada kelas positif yang tidak didaftarkan sebagai syarat positif [6][7].

5. Kesimpulan

Sistem dan penelitian ini dibangun untuk mengklasifikasikan data *Twitter* dalam memprediksi sebuah sentimen yang akan menghasilkan akurasi pada setiap pengujian menggunakan pembobotan TF-IDF pada metode *Naive Bayes Classifier.* Dari hasil yang telah didapatkan, dapat disimpulkan bahwa hasil klasifikasi sentimen terhadap kandidat calon Presiden pada Pilpres 2019 berdasarkan hasil dari data yang diujikan adalah sentimen yang bersifat positif.

Penggunaan metode pembobotan TF-IDF pada *Naive Bayes Classifier* menghasilkan nilai akurasi 57,00%. Nilai tersebut dihasilkan dari proses TF-IDF menggunakan bentuk kata unigram, bigram, dan trigram. Nilai akurasi tersebut belum dapat dikatakan berhasil karena jumlah data yang telah *crawled* hanya sebesar 500 data *Twitter* dan masih banyak faktor lain yang menyebabkan kecilnya akurasi yang didapatkan. Hal ini dapat dikembangkan atau ditingkatkan dengan penambahan fitur yang lebih lengkap juga *word sentiment* yang lebih jelas dan detail [14].

Daftar Pustaka

[1] Casanova, M., Breitman, K., & Truszkowski, W. (2007). Semantic Web: Concepts, Technologies and Applications. (3), 155-173.

[2] Davies, J., Fensel, D., & Harmelen, F. v. (2003). *Towards The Semantic Web Ontology-driven Knowledge Management.* John Wiley & Sons, Ltd.

[3] Garcia, S. J. (2015). *Data preprocessing in data mining.* Switzerland: Springer.

[4] Haewoon Kwak, C. L. (2010). What is Twitter, a social network or a news media?

[5] Lan, M. e. (2009). Supervised and traditional term weighting methods for automatic text categorization. *Pattern Analysis and Machine Intelligence*, 721-735.

[6] Liu, B. (2012). *Sentiment Analysis and Opinion Mining.* Morgan & Claypool Publishers.

[7] Markov, Z., & Russell, I. (n.d.). *An Introduction to the WEKA Data Mining System.* Central Connecticut State University & University of Hartford.

[8] P, A. R. (2011). *Sentiment Classification for Indonesian Message in Social Media.* Bandung: Bandung Institute of Tehcnology Conference.

[9] S, B. F. (2014). *[Pengenalan] Apa Itu Twitter API dan Pembuatan Consumer Key dan Consumer Secret ?* Retrieved Januari 2016, 21, from

[10] Tokunaga, T. I. (1994). Text categorization based on weighted inverse document frequency. *Special Interest Groups and Information Process Society of Japan (SIG-IPSJ*.

[11] Sembodo, J. E., Setiawan, E. B., & Baizal, Z. A. (2016). Data Crawling Otomatis pada Twitter. In *Indonesian Symposium on Computing (Indo-SC)* (pp. 11-16)

[12] Saraswati, N.W.S., 2011, Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machine untuk Sentimen Analysis.

[13] Noviah Dwi Putranti, Edi Winarko (2014).”Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine”.

[14] Go, A., Huang, L., & Bhayani, R. (2009). Twitter Sentiment Analysis. Final Project Report, Stanford University, Department of Computer Science.

[15] Widodo, A.W., 2013. Klasifikasi Artikel Berita Menggunakan Naive Bayes Classifier yang Dimodifikasi.