JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VII No. 1, Dec 2020, hlm. 99 – 104 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v7i1.851
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

ANALISIS SENTIMEN PADA PEMERINTAHAN TERPILIH PADA PILPRES 2019 DI*TWITTER* MENGGUNAKAN ALGORITME *NAÏVEBAYES*

Febby Apri Wenando[1*], Regiolina Hayami[1], Agung Jefrianto Anggrawan[1]
[1]Teknik Informatika, Universitas Muhammadiyah Riau
email: [*]febbyapri@umri.ac.id

Abstract: The Presidential general election on 2019 became one of the most popular topics on twitter nowdays. The society give their opinion about the pair of candidates that they are support through the social media. This research was predicts about the society sentimens toward the candidates of President and Vice President of Republic of Indonesia. The data was used based on the tweet on the @jokowi twitter account. The retrieval of data by using the Tweepy library with the Python 2.7 programming language. This research was classified became of two of society sentiments classes, namely positive and negative. The modeling was used of the weighting method Unigram, Bigram, Trigram, N-Gram (1-2) and N-Gram (1-3) that used the Naïve Bayes Algorithm on the Weka Application. The modeling data was used by the dataset of 646 sentences. The highest results of this reseach were obtained by Unigram Weighting, namely: 81.4% accuracy, 81.5% precision, 81.3% recall with a time of 0.3 s.

Keywords: classification, naïve bayes, 2019 presidential election, twitter, unigram

Abstrak: Pemilihan Umum tentang Pilpres 2019 menjadi salah satu topik yang ramai diperbincangkan di *Twitter*. Adu pendapat di sosial media oleh masyarakat mengandung opini terhadap pasangan calon yang didukungnya. Penelitian ini memprediksi sentimen masyarakat kepada pasangan calon Presiden dan Wakil Presiden Republik Indonesia. Data yang digunakan adalah tweet yang ada pada akun Twitter @jokowi. Pengambilan data menggunakan *library* Tweepy dengan bahasa pemrograman Python 2.7. Penelitian ini mengklasifikasi sentimen masyarakat menjadi 2 kelas, yaitu positif dan negatif. Kemudian dilakukan pemodelan dengan metode pembobotan Unigram, Bigram, Trigram, N-Gram (1-2) Dan N-Gram (1-3) menggunakan Algoritme Naïve Bayes pada Aplikasi Weka. Pembuatan model menggunakan dataset yang berjumlah 646 kalimat. Hasil tertinggi yang diperoleh pada penelitian ini adalah dengan menggunakan Pembobotan Unigram, yaitu : akurasi 81,4%, presisi 81,5 % , recall 81,3 % dengan catatan waktu 0,3s.

Kata kunci: klasifikasi, naïve bayes, pilpres 2019, twitter, unigram.

PENDAHULUAN

Indonesia merupakan negara yang menganut sistem demokrasi. Hal tersebut dibuktikan dengan diada-kannya pemilihan umum dalam pemerintah baik pusat maupun daerah. Pemilu diselenggarakan dalam 5 tahun sekali. Perbincangan terkait Pemilu ini menjadi hangat diakalangan masya-rakat, berbagai pikiran dan padangan politik dalam PILPRES baik yang pro maupun kontra sudah semakin banyak bermunculan, misalnya munculnya *hastag* #gantipresiden dan lain sebagainya.

Twitter termasuk media sosial yang ramai digunakan masyarakat saat ini untuk saling berbagi dan bertukar

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VII No. 1, Dec 2020, hlm. 99 – 104 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v7i1.851
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

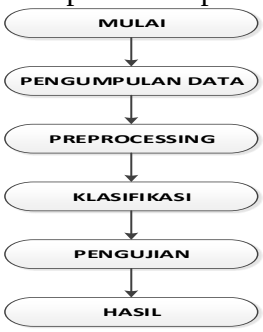
informasi. Saat ini *Twitter* banyak digunakan untuk berbagai kepentingan, seperti media sosial untuk membangun pertemanan dalam skala luas dan untuk saling bertukar fikiran dan menyampaikan pendapat. Selain itu, *Twitter* juga dapat digunakan untuk alat promosi dan kampanye dalam hal kepentingan politik. Data di media sosial twitter yang berupa teks, maka data ini dapat di olah menjadi sebuah informasi dan pengetahuan yang berguna terhadap isu yang berkembang[1].

Text mining merupakan teknik penambangan data teks yang bertujuan untuk mendapatkan kembali informasi yang ada pada data teks, yang diekstrak secara otomatis dari sumber-sumber data teks yang digunakan sebagai dataset[1] [2]. Pada penelitian terkait analisis sentimen, digunakan *dataset* dari pendapat atau opini masyarakat[3].

Selanjutnya pendapat/ opini masyarakat tersebut dibagi menjadi 5(lima) kelas. Untuk dapat meng-hasilkan suatu klasifikasi, penelitian ini menggunakan salahsatu jenis metode klasifikasi yaitu *Naïve Bayes Classifier*(NBC). Dataset yang digu-nakana adalah data teks bahasa Indonesia berupa *tweet* dari *Twitter* tentang Calon Presiden Indonesia tahun 2014. Hasil yang diperoleh dari penggunaan 900 dataset pada penelitian tersebut yaitu nilai akurasi sebesar 71,9%, nilai presisi sebesar 71,6%, dan nilai recall sebesar 71,9%. Sebuah model dibangun untuk melakukan analisis pada Pemilihan Gubernur dan Wakil Gubernur DKI Jakarta tahun 2017 [4] untuk mengetahui sentimen masyarakat di Twitter terhadap pasangan calon Gubernur dan Wakil Gubernur. Data yang digunakan diperoleh dengan menggunakan kata kunci(*keyword)* @AhokDjarot dan @JktMajubersama. Hasil penelitian tersebut adalah prediksi data uji menggunakan algoritme *Naïve Bayes*, dengan tingkat akurasi mencapai 60,60%.

METODE

Tahapan dalam menyelesaikan penelitian ini dapat dilihat pada Gambar 1.



Gambar 1 Metodologi Penelitian

Pengumpulan Data

Penelitian ini menggunakan dataset Sentimen masyarakat PILPRES yang akan diambil dari data berisi *tweet* yang telah dianotasikan untuk PILPRES, khususnya dalam kategori pro dan kontra pada akun *@jokowi*. Kemudian data dibagi menjadi 2 kelas yaitu kelas positif dan kelas negatif.

Preprocessing

Preprocessing (pra-proses) pada penelitian ini bertujuan untuk menghapus data-data *tweet error* pada saat pengambilan data dan menyeleksi fitur berupa kata-kata atau *term*. Praproses penelitian ini menggunakan aplikasi Python ver. 2.7. Tahapan *preprocessing*, terdiri dari proses pembersihan karakter[4], *stemming*[5] dan *stopword*[6].

Klasifikasi

Data teks yang telah dilakukan pembersihan pada tahap *preprocesing*. Data teks tersebut akan diberikan bobot

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VII No. 1, Dec 2020, hlm. 99 – 104 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v7i1.851
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

menggunakan Metode Pembobotan Kata TF-IDF N-Gram (1-2) dan N-Gram (1-3) yang bertujuan untuk mempresentasikan seberapa besar pengaruh bobot tersebut pada suatu dokumen[7]. Cara kerja pembobotan kata pada metode TF-IDF yaitu dengan menggunakan 2 parameter pembo-botan yaitu pembobotan lokal *tfi.j* adalah bobot yang didapat dari frekuensi kemunculan kata *i* dalam dukumen *j* dan pembobotan global dengan menggunakan *idfi* adalah bobot yang didapat dengan memper-timbangkan jumlah kemunculan kata *i* (DF*i*) pada keseluruhan dukumen N. Dan selanjutnya, nilai dari bobot lokal dikalikan dengan nilai dari bobot global maka didapat[8]. Cara menghitung bobotnya dengan persamaan(1) [7]:

$$w_{i,j} = tf_{i,j}\ x\ (\log\left(\frac{N}{DF_i}\right)) \quad (1)$$

Setelah *dataset* dibobotkan kemudian dilakukan klasifikasi menggunakan algoritme *Naïve Bayes*[1]. Algoritme Naïve Bayes menggunakan pendekatan Bayes dalam melakukan proses klasisfikasi. Untuk mencari nilai probabalitias/ peluang tertinggi (Vmap) digunakan rumus(2) berikut: [8]

$$Vmap = arg\ P(Vj|a1, a2, a3, \ldots, an) \quad (2)$$

Keterangan:
Vmap=Probabilitas/ peluang tertinggi
a1, a2, a3, .., an =atribut data masukan
Persamaan diatas dapat ditulis menjadi rumus (3): [8]

$$Vmap = \arg max \frac{p(a1, a2, a3, \ldots, an|Vj)|p(Vj)}{p(a1, a2, a3, \ldots, an)} \quad (3)$$

Keterangan:
Vmap = Probabilitas
P(Vj) = Probabilitas kelas ke *j*
P(a1,a2,…,an) = Probabilitas atribut input
P(a1,a2,..,an|Vj) = Probabilitas atribut input jika diketahui keadaan V*j* ke *j*

Pada penelitian yang dilakukan, pengolahan dataset menggunakan algoritme *Naïve Bayes* akan diimplementasikan pada suatu sistem data mining yang disebut WEKA. Sistem ini berisikan *tools* yang mengimplementasikan algoritme *data mining*. WEKA digunakan untuk pra-pengolahan data (*preprocessing*), klasi-fikasi data, regresi, klasterisasi (*clustering*), aturan asosiasi dan visualisasi data[9].

HASIL DAN PEMBAHASAN

Pengumpulan data

Proses pengambilan data menggunakan Python 2.7, data yang digunakan adalah berupa teks berisikan sentimen masyarakat pada akun *twitter* @jokowi dengan format CSV yang berbentuk data tidak terstruktur[9], kemudian untuk mempermudah klasifikasi, peneliti memberikan label yang berfungsi untuk menyimpan nilai sentimen yang akan diklasifikasikan [10]. Data yang dikumpulkan dalam penelitian ini berjumlah 646 *tweet*, 476 teks yang bersifat positif dan 170 yang bersifat negatif. *Dataset* ini dikategorikan kepada *unbalanced class.*

Preprocessing

Preprocessing ini dilakukan dengan tiga tahap, yaitu:

1. Pembersihan karakter, yaitu menghilangkan karakter yang bisa mengurangi kualitas dataset yang akan dilatih. Seperti tanda *hastag* (#), *at* (@) dan simbol karakter lainnya yang dianggap tidak perlu, seperti tercantum pada Tabel 1 .
2. *Stemming* adalah proses suatu kata menjadi kata dasar. Pada tahap ini akan dilakukan penghilangan semua imbuhan (*affix*) yang terdiri dari awalan (*prefix*), sisipan (*infix*), akhiran (*suffix*) dan duplikasi[6]. Tujuan dari proses ini adalah untuk

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VII No. 1, Dec 2020, hlm. 99 – 104 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v7i1.851
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

mendapatkan kata dasar dari kata berimbuhan.

3. *Stopword*, adalah proses pembuangan kata-kata yang tidak memiliki arti atau tidak relevan. Kata-kata yang diperoleh dari tahap tokenisasi akan dicek pada daftar *stopword*, apabila sebuah kata tersebut masuk dalam daftar *stopword* maka kata tersebut akan dihilangkan dan tidak akan diproses lebih lanjut. Sebaliknya apabila sebuah kata tidak termasuk dalam daftar *stopword* maka kata tersebut akan masuk ke proses berikutnya.

Tabel 2 dan 3 merupakan contoh penerapan tahap 2 dan 3 dari *preprocessing* terhadap *dataset*.

Klasifikasi

Data bersih yang didapat setelah tahap *preprocessing*, kemudian dilakukan pemisahan tiap kata dengan menggunakan Aplikasi WEKA[9]. Pada proses klasifikasi data masukan yang digunakan adalah dokumen dengan tipe *Atribut-Relation File Format*(ARFF). Hasil dari pemisahan tersebut selanjutnya dibobotkan dan diklasifikasi menggunakan algoritme *Naïve Bayes*.

Pengujian

Pada tahapan pengujian hasil klasifikasi yang diperoleh selanjutya diuji. Pengujian hasil klasifikasi menggunakn teknik *K-Fold Cross Validation*, yaitu dengan jumlah percobaan (k) sebanyak 10 kali. Pengujian dilakukan dengan membagi *dataset* mejadi 10 bagian. Sebanyak 9(Sembilan) dari 10(sepuluh) bagian *dataset* digunakan untuk proses pelatihan(*training)*. Sisanya digunakan untuk proses pengujian(*testing*). Iterasi atau perulangan terjadi sebanyak 10(sepuluh) kali dengan variasi kombinasi data sebanyak 10(sepuluh) bagian pada *dataset* untuk *training* maupun *testing*.

Tahapan pengujian dilakukan untuk menganalisa hasil dari pembelajaran mesin yang telah dilakukan, sehingga mendapatkan hasil akurasi, presisi, *recall*, dan waktu klasifikasi.

Tabel 1 Pembersihan Karakter

Label	Teks Awal	Setelah Proses Pembersihan
Positif	RT @jokowi: telah berjuang untuk mengharumkan nama bangsa Indonesia di mata dunia	RT jokowi telah berjuang untuk mengharumkan nama bangsa Indonesia di mata dunia
Positif	RT @roninpribumi: Kata @jokowi demokrasi ada batasnya itu betul. Batasnya adalah HUKUM. Kalau tagar #2019GantiPresiden dianggap melanggar…	RT roninpribumi Kata jokowi demokrasi ada batasnya itu betul Batasnya adalah HUKUM Kalau tagar GantiPresiden dianggap melanggar…
Negatif	RT @eae18: Saya bingung deh. Kan ada Yai Ma'ruf. Kan ada @imam_nahrawi Mosok Pak @jokowi kutip hadis yang salah konteks didiamkan.	RT eae Saya bingung deh Kan ada Yai Maruf Kan ada imam nahrawi Mosok Pak jokowi kutip hadis yang salah konteks didiamkan

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VII No. 1, Dec 2020, hlm. 99 – 104 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v7i1.851
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Tabel 2 *Stemming*

Label	Teks Awal	Setelah Proses *Stemming*
Positif	RT @jokowi: telah berjuang untuk mengharumkan nama bangsa Indonesia di mata dunia	jokowi telah juang untuk harum nama bangsa Indonesia di mata dunia
Positif	RT @roninpribumi: Kata @jokowi demokrasi ada batasnya itu betul. Batasnya adalah HUKUM. Kalau tagar #2019GantiPresiden dianggap melanggar…	roninpribumi kata jokowi demokrasi ada batas itu betul batas adalah hukum kalau tagar gantipresiden anggap langgar
Negatif	RT @eae18: Saya bingung deh. Kan ada Yai Ma'ruf. Kan ada @imam_nahrawi Mosok Pak @jokowi kutip hadis yang salah konteks didiamkan.	eae saya bingung deh kan ada yai maruf kan ada imam nahrawi mosok pak jokowi kutip hadis yang salah konteks diam

Tabel 3 *Stopword*

Label	Teks Awal	Setelah Proses *Stopword*
Positif	RT @jokowi: telah berjuang untuk mengharumkan nama bangsa Indonesia di mata dunia	Jokowi berjuang harum nama bangsa Indonesia mata dunia
Positif	RT @roninpribumi: Kata @jokowi demokrasi ada batasnya itu betul. Batasnya adalah HUKUM. Kalau tagar #2019GantiPresiden dianggap melanggar…	roninpribumi kata jokowi demokrasi batas betul batas hukum kalau tagar gantipresiden anggap langgar…
Negatif	RT @eae18: Saya bingung deh. Kan ada Yai Ma'ruf. Kan ada @imam_nahrawi Mosok Pak @jokowi kutip hadis yang salah konteks didiamkan.	eae bingung deh kan yai ma ruf kan imam nahrawi mosok pak jokowi kutip hadis salah konteks diam

SIMPULAN

Pada Klasifikasi ini peneliti menggunakan Algoritme Naive Bayes dengam pembobotan N-gram (1-2) dan N-gram(1-3). Hasilnya dapat dilihat pada tabel 4:

Tabel 4. Hasil Klasifikasi

Naïve Bayes	**N-gram (1-2)**	**N-gram (1-3)**
Accuracy (%)	80,5	80,2
Precision (%)	80,7	80,2
Recall (%)	80,3	80,2
Time (s)	0,42	0,43

Berdasarkan Tabel 4 dapat dilihat perbandingan persentase hasil akurasi(*accuracy*), presisi(*precision*), *recall* serta waktu(*time*). Dari hasil penggujian dapat dilihat bahwa algoritme *Naïve Bayes* dengan TF-IDF berbasis N-Gram (1-2) menunjukkan hasil yang paling baik, dalam hal nilai akurasi(*accuracy*), presisi(*precision*), *recall* serta waktu(*time*) klasifikasi. Dengan nilai akurasi(*accuracy*) sebesar 80,5%, presisi(*precision*) sebesar 80,7%, *recall* sebesar 80,3% dengan catatan waktu(*time*) selama 0,42s. Namun, hasil

tersebut tidak jauh berbeda dengan N-Gram (1-3). Melalui penelitian ini juga diketahui bahwa Metode Pembobotan Kata N-Gram (1-2) dan (1-3) hasilnya sama baik dan tidak jauh berbeda ketika dilakukan.

UCAPAN TERIMA KASIH

Terima kasih kepada Kementerian Riset, Teknologi dan Pendidikan Tinggi(Kemenristek-Dikti) atas dukungan pendanaan dalam penelitian ini.

DAFTAR PUSTAKA

[1] & S. J. Feldman, R, *Text Minning Handbook : Advanced Approaches in Analyzing Unstructure Data*. New York: Cambridge University Press., 2007.

[2] M. A. Razzaq, A. M. Qamar, and Hafiz Syed Muhammad Bilal, "Prediction and analysis of Pakistan election 2013 based on sentiment analysis," in 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Aug. 2014, pp. 700–703, doi: 10.1109/ASONAM.2014.6921662.

[3] G. A. Buntoro, T. B. Adji, and A. E. Purnamasari, "Sentiment Analysis Twitter dengan Kombinasi Lexion Based dan Double Propagation," *CITEE*, pp. 39–42, 2014.

[4] M. H. Rasyadi, "Analisis Sentimen Pada Twitter Menggunakan Metode Naïve Bayes(Studi Kasus Pemilihan Gubernur DKI Jakarta 2017)," 2017.

[5] F. Nausheen and S. H. Begum, "Sentiment analysis to predict election results using Python," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, Jan. 2018, pp. 1259–1262, doi: 10.1109/ICISC.2018.8399007.

[6] F. . Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," Inst. Log Lang. Comput. Univ. Van Amst. Neth., 2003.

[7] F. . Wenando, T. B. Adji, and Ardiyanto, "Text Classification to Detect Student Level of Understanding in Prior Knowledge Activation Process," *Adv. Sci. Lett.*, vol. 23, no. 3, pp. 2285–2287, 2017.

[8] F. . Wenando and E. Fuad, "Detection of Hate Speech in Indonesian Language on Twitter Using Machine Learning Algorithm," *Pros. CELSciTech*, vol. 4, pp. 6–8, 2019.

[9] S. S. Aksenova, *Mechine Learning with WEKA – WEKA Tutorial – Explore Tutorial for WEKA Version 3.4.3*. California, 2004.

[10] K. D. A, Normah, and U. A. H, "Prediction of Indonesia Presidential Election Results for the 2019-2024 Period Using Twitter Sentiment Analysis," *2019 5th Int. Conf. New Media Stud.*, pp. 36–42, 2019.