

IMPLEMENTASI TEXT MINING UNTUK ANALISIS OPINI PUBLIK TERHADAP CALON PRESIDEN

Trya Sovi Kartikasari¹⁾, Hendry Setiawan²⁾, Paulus Lucky Tirma Irawan³⁾

^{1,2,3}Prodi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Ma Chung Jl.
Villa Puncak Tidar N-01, Malang

E-mail: ¹tryasovi@gmail.com, ²hendry.setiawan@machung.ac.id,

³paulus.lucky@machung.ac.id

ABSTRAK

Sistem presidensial merupakan salah satu bentuk demokrasi di Indonesia. Sistem tersebut menitikberatkan pada penyelenggaraan pemilihan umum presiden dan wakilnya yang dilakukan secara langsung oleh rakyat. Tingkat terpilihnya seorang presiden dapat dilihat dari opini publik yang beredar, salah satunya pada media sosial yang juga merupakan bagian dari kampanye. Dalam penelitian ini akan dianalisa opini yang berkaitan dengan elektabilitas calon presiden dari media sosial Twitter dari media sosial Twitter menggunakan metode Naïve Bayes Classifier (NBC) dan menentukan faktor-faktor yang terbentuk dari opini menggunakan Principal Component Analysis (PCA). Data opini dari media sosial Twitter didapatkan menggunakan kata kunci "Jokowi" dan "Prabowo". Sebagian opini tersebut dipilih sebagai data latih untuk didapatkan kelas bersentimen negatif dan positif. Setelah proses pelatihan, dilakukan proses terhadap data uji dan data validasi. Hasil akurasi untuk data uji topik Jokowi pada tweet bersentimen positif mendapatkan akurasi sebesar 88.63% dan negatif sebesar 91.06%. Sementara untuk Prabowo bersentimen positif mendapatkan akurasi sebesar 88.58% dan negatif sebesar 80.37%. Rerata akurasi untuk keseluruhan topik adalah adalah 86.89%. Untuk mendapatkan faktor pada setiap sentimen, dilakukan proses perhitungan nilai PCA. Setiap sentimen tersebut kemudian dilakukan analisis faktor oleh pakar, yakni didapatkan 20 faktor yang sudah berhasil diinterpretasikan oleh pakar.

Kata Kunci : *Naive Bayes Classifier, Analisis Sentimen, Principal Component Analysis, Text Mining*

ABSTRACT

Presidential system is one form of democracy in Indonesia. The system emphasizes the implementation of president and vice president election that conducted directly by peoples. The level of President election can be seen from public opinion which is also part of the campaign. In this research, opinion will be analyzed relate to the electability of presidential candidate from media social Twitter using Naïve Bayes Classifier (NBC) method and determining the factors formed form opinion using Principal Component Analysis (PCA). The opinion data from social media Twitter is obtained using the keywords "Jokowi" and "Prabowo". Some of the opinions were chosen as training data to obtain negative and positive sentiment classes. After the training process, a process is carried out on the test data and validation data. The accuracy results for Jokowi's topic test data on positive-sent tweets get an accuracy of 88.63% and negative at 91.06%. While for Prabowo positive sentiment get an accuracy of 88.58% and negative at 80.37%. The total accuracy of each topic is 86.89%. To get factors in each sentiment, PCA value is calculated. Each of these sentiments is carried out by a factor analysis by experts, which is obtained 20 factors that have been successfully interpreted by experts.

Keywords : *Naive Bayes Classifier, Sentiment Analysis, Principal Component Analysis, Text Mining..*

PENDAHULUAN

Media sosial saat ini menjadi *platform* untuk berbagi bermacam-macam informasi. Dari laporan yang diterbitkan pada tanggal 30 Januari 2018 yang berjudul “*Essential Insight Into Internet, Social Media, Mobile and E-commerce Use Around The World*”, total populasi di Indonesia sebanyak 265,4 juta jiwa dengan pengguna aktif media sosial mencapai 130 juta. Media yang paling banyak digunakan, yakni Youtube, Facebook, Whatsapp, Instagram, Line, BBM, Twitter, Google+, Fb Messenger, LinkedIn, Skype dan Wechat. Saat ini telah terjadi gencar-gencarnya informasi tentang pemilihan presiden tahun 2019, informasi tersebut beredar luas pada media sosial khususnya Twitter. Selain melakukan kampanye secara langsung, calon presiden dan para pendukungnya pun melakukan kampanye pada Twitter. Informasi ataupun kampanye tersebut memberikan *feedback* berupa opini dari masyarakat baik berupa positif maupun negatif.

Beberapa metode sudah cukup untuk menganalisis suatu opini pada *platform* media sosial yang lain. Pada penelitian sebelumnya telah dilakukan penerapan analisis opini publik pada media sosial Twitter menggunakan metode Naïve Bayes Classifier (NBC) dan menunjukkan hasil nilai akurasi [1]. Penelitian serupa juga telah dilakukan yakni penerapan analisis sentimen akun resmi tokoh publik dalam media sosial Twitter menggunakan metode NBC dan disimpan ke dalam bentuk *database MySQL* yang memperoleh hasil akurasi analisis sentimen positif, negatif dan netral [2].

Dalam penelitian ini akan dilakukan penerapan metode *Text Mining* untuk menganalisa sumber data opini yang cukup banyak di media sosial Twitter dan membangun sistem analisis opini publik. Tujuan sistem analisis ini adalah sebagai media untuk analisis opini

publik terhadap elektabilitas calon presiden dan faktor-faktor yang mempengaruhi elektabilitas tersebut agar menghemat tenaga dan waktu karena pergerakan data dari media sosial yang sangat cepat. Dalam penelitian ini, terdapat batasan masalah, yakni pengambilan data *tweet* diolah setelah debat I dan II Cawapres, menggunakan bahasa pemrograman Python, menggunakan *web* untuk mengimplementasikan sistem analisis sentimen dan faktor, analisis menggunakan metode NBC, ekstraksi variabel penentu menggunakan *Principal Component Analysis* (PCA) dan terdapat admin untuk validasi klasifikasi sentimen sampai tahap *Document Term Matrix* (DTM).

METODE

Penelitian dilakukan dalam 6 tahapan utama. Tahap pertama, yakni proses pengumpulan data tweets melalui Twitter API dimana peneliti mendapatkan Keys dan tokens untuk otentifikasi untuk mengunduh informasi pada Twitter. Tahap kedua, yakni text preprocessing, metode yang digunakan adalah case folding, tokenizing, filtering dan stemming. Tahap ketiga, yakni DTM. DTM merupakan hasil dari preprocessing yang terdiri atas normalisasi teks (case folding, tokenizing, filtering, stemming) atau sebuah matriks yang menggambarkan frekuensi kata dalam dokumen.

Tahap keempat, yakni seleksi fitur. Seleksi fitur (feature selection) didasarkan pada eliminasi suatu atribut yang kurang relevan [3]. Tahap kelima, yakni analisis sentimen menggunakan metode NBC. NBC merupakan metode klasifikasi sentimen yang berdasar pada probabilitas dan Teorema Bayesian yang memiliki asumsi bahwa setiap variabel X merupakan variabel yang bersifat bebas [4]. NBC adalah metode yang paling sederhana, metode ini memiliki akurasi serta performansi yang tinggi dalam klasifikasi teks [5]. Pada tahap terakhir

dilakukan analisis faktor. Analisis faktor adalah teknik analisis memuat tentang ketergantungan beberapa variabel secara simultan dengan tujuan penyederhanaan bentuk hubungan antara beberapa variabel yang telah diteliti menjadi sejumlah faktor yang berjumlah sedikit daripada variabel yang sudah diteliti [6].

Tahapan Pengumpulan Data

Application Programming Interface (API) merupakan cara program komputer untuk dapat berbicara satu sama lain agar informasi dapat diminta dan disajikan kepada pengguna. Twitter API didapatkan dengan cara mendaftarkan akun Twitter pada laman <https://developer.twitter.com/>. Secara default, hanya informasi publik yang dapat diakses pada aplikasi yang dibuat di Twitter. Setelah akun berhasil didaftarkan, maka *consumer key*, *consumer access*, *access token* dan *access token secret* akan disediakan seperti Gambar 1 berikut.



Gambar 1 Twitter API

Twitter API berfungsi sebagai syarat otentifikasi dari aplikasi yang dibangun sebelum dapat mengunduh data yang ada pada web Twitter. Setelah keys dan tokens Twitter didapatkan, maka kode tersebut wajib dimasukkan ke dalam aplikasi yang akan dibuat untuk memberikan akses untuk mengunduh data informasi dari Twitter.

Tahapan Preprocessing

Text mining dilakukan setelah teks dalam dokumen yang digunakan sudah dipersiapkan terlebih dahulu dan masuk ke dalam proses utama. Tahapan proses teks dalam dokumen atau data mentah dipersiapkan dinamakan *text processing*.

Data mentah memiliki beberapa karakteristik seperti gangguan (noise) dan teks yang tidak terstruktur. Data dalam teks yang tidak terstruktur tersebut diubah menjadi data teks yang terstruktur. Berikut merupakan tahap preprocessing.

1. Case Folding

Penggunaan huruf kapital teks dalam text processing harus konsisten. Fungsi dari case folding adalah untuk penyamaan case dalam teks. Oleh karena itu, keseluruhan teks dalam dokumen akan diproses dan dikonversi menjadi bentuk standar. Bentuk standar tersebut berupa huruf kecil.

2. Tokenizing

Tahapan tokenizing merupakan proses pemotongan dokumen string input pada tiap penyusun kata yang berasal dari hasil case folding.

3. Filtering

Filtering merupakan proses seleksi atau pengambilan kata yang dihasilkan oleh proses tokenizing dan dilakukan dengan menggunakan cara stopword maupun wordlist

4. Stemming

Stemming merupakan proses penemuan akar kata dari hasil filtering dengan menggunakan aturan tertentu.

Document Term Matrix (DTM)

Hasil dari preprocessing yang terdiri atas normalisasi teks (case folding, tokenizing, filtering, stemming) disajikan dalam bentuk Document Term Matrix (DTM). DTM adalah matriks yang menggambarkan frekuensi kata dalam dokumen. Dokumen ini berisi kumpulan dari tweet. Seperti pada Tabel 1, jumlah kolom dalam DTM menunjukkan banyaknya kata yang didapatkan dari tahap preprocessing, sedangkan jumlah

baris tweet menggambarkan banyaknya tweet.

Angka 7 pada kolom Kata 1 dan baris Tweet 3 berarti bahwa kata ke-1 memiliki frekuensi sebanyak 7 pada dokumen Tweet 3 dan seterusnya.

Tabel 1. DTM

No	Kata 1	Kata 2	Kata 3	Kata 4	Kata 5
Tweet1	10	5	16	5	2
Tweet2	4	9	10	20	15
Tweet3	7	9	3	2	8
Tweet4	0	10	3	5	7
Tweet5

Seleksi Fitur

Dimensi dari sebuah kumpulan teks dikurangi dengan menghapus kata-kata yang tidak menggambarkan isi dari dokumen sehingga proses klasifikasi lebih akurat. Dari hasil DTM tersebut, kata-kata yang dikurangi adalah yang memiliki frekuensi yang rendah dengan tujuan untuk meningkatkan akurasi dari klasifikasi yang dilakukan. Kata 4 pada Tabel 2 memiliki frekuensi sebanyak 25, berarti bahwa memiliki frekuensi yang rendah daripada kata 1, kata 2, kata 3, dan lain sebagainya. Akibatnya, kata 4 tersebut dihilangkan.

Tabel 2. Tabel Frekuensi

No	Kata	Frekuensi (f)
1	Kata 1	38
2	Kata 2	58
3	Kata 3	29
4
5	Kata

Analisis Sentimen Naïve Bayes Classifier

Ada dua tahap dalam klasifikasi teks pada metode Naïve Bayes Classifier, yaitu tahap pembelajaran dan tahap klasifikasi.

1. Proses Pembelajaran Naïve Bayes Classifier

Pada proses pembelajaran, data latih digunakan dan dilakukan perhitungan nilai ln probabilitas untuk setiap term yang ada. Rumus probabilitas dapat dilihat pada (1).

$$p(w|Ck) = \frac{f+1}{\sum w(fw+1)} \tag{1}$$

Kasus atau penelitian yang memiliki data dalam jumlah banyak, tidak dapat teratasi dengan persamaan tersebut. Karena persamaan ini memiliki nilai probabilitas yang sangat kecil dan keterbacaan oleh sistem terganggu Sehingga dilakukan proses normalisasi nilai dengan menggunakan fungsi log natural (ln) dan persamaan dapat dilihat pada (2).

$$\ln p(w|Ck) = \ln \frac{f+1}{\sum w(fw+1)} \tag{2}$$

Dimana:

- Ck : kelas positif atau negatif
- w : kata dalam frekuensi
- f : frekuensi kata yang muncul
- fw : frekuensi kata data latih
- 1 : additive smoothing (menghindari frekuensi kata nol)
- 2. Proses Klasifikasi Naïve Bayes Classifier

Tahap terakhir dalam proses klasifikasi ini adalah menghitung probabilitas setiap kelas pada tweet. Hasil klasifikasi yang ditentukan adalah dari perhitungan probabilitas yang memiliki nilai terbesar. Persamaan probabilitas klasifikasi ini dapat dilihat pada (3).

$$\ln p(Ck|t) = \sum_y \ln P_y + \sum_x f_x * \ln \frac{1}{\sum w(fw+1)} \tag{3}$$

Dimana:

- Ck : kelas positif atau negatif
- t : tweet
- w : kata data latih
- y : kata data uji yang ada pada data latih
- x : kata pada data uji yang tidak ada pada data latih
- lnPy : Ln probabilitas kata yang terdapat pada data latih
- fx : frekuensi kata data uji yang tidak ada pada data latih
- fw : frekuensi kata data latih

Analisis Faktor

Analisis faktor digunakan untuk mengidentifikasi faktor yang memiliki jumlah relatif kecil yang dapat digunakan untuk menjelaskan variabel yang saling berhubungan dalam jumlah yang banyak. Variabel-variabel dalam satu faktor tersebut memiliki korelasi yang tinggi, sedangkan variabel-variabel yang berada pada faktor lain memiliki korelasi yang relatif rendah. Berikut merupakan tiga fungsi analisis faktor, yaitu [7]:

1. Dimensi-dimensi yang mendasar diidentifikasi dan dapat dijelaskan sebagai korelasi dari serangkaian variabel.
2. Variabel-variabel baru yang kecil diidentifikasi untuk menggantikan variabel yang tidak berkorelasi dari serangkaian variabel asli yang berkorelasi.
3. Beberapa variabel kecil diidentifikasi dari sejumlah variabel yang banyak untuk dianalisis multivariate lainnya.

Metode analisis faktor yang dipakai pada penelitian ini adalah PCA. PCA merupakan teknik analisis faktor dimana beberapa faktor yang akan terbentuk adalah berupa variabel yang belum dapat ditentukan sebelum analisis dilakukan. Gambaran umum PCA dapat dilihat pada Gambar 2.



Gambar 2. PCA

Metode PCA digunakan untuk masing-masing sentimen positif dan negatif dalam bentuk frekuensi kata. Masing-masing frekuensi kata dilakukan perhitungan vektor rata-rata. Setelah vektor rata-rata didapatkan, dilakukan perhitungan untuk selisih matriks dari vektor asal dan vektor rata-rata. Dari hasil selisih matriks tersebut, dihitung

matrix kovarian dengan persamaan (4).

$$A = Y^T x Y \quad (4)$$

Dimana:

A : Matriks kovarian

Y^T : Selisih matriks vektor asal dan vektor rata-rata (transpos)

Y : Selisih matriks vektor asal dan vektor rata-rata

Setelah menghitung matriks kovarian, maka langkah selanjutnya mencari nilai eigen dan vektor eigen dari hasil matriks kovarian tersebut. Hasil dari nilai eigen tersebut diurutkan dari hasil paling besar hingga kecil. Langkah selanjutnya adalah melakukan perhitungan nilai PC yang akan digunakan sebagai faktor penentu dalam pengelompokannya dan perhitungan tersebut dapat dilihat pada persamaan (5).

$$PC = (Yxv) \quad (5)$$

Dimana:

PC : Principal Component

Y : Selisih matriks vektor asal dan vektor rata-rata

v : Nilai eigen

Pengembangan Program

1. Pengambilan Data dan Text Mining
Pengguna melakukan pengambilan data yang digunakan sebagai data uji melalui Twitter API dan diperoleh data tweet. Data disimpan ke dalam bentuk file CSV. Data tersebut dilakukan proses preprocessing oleh sistem yang terdiri dari tahap case folding, tokenizing, filtering dan stemming dengan menggunakan library yang disediakan oleh Python. Setelah tahap preprocessing selesai, maka dilakukan tahap seleksi fitur dengan menghilangkan kata yang memiliki jumlah huruf kurang dari tiga.
2. Analisis Sentimen

Setelah tahap pengambilan data dan proses Text Mining (preprocessing dan seleksi fitur) selesai, maka dilakukan klasifikasi tweet terhadap data uji dengan menggunakan persamaan dalam metode Naïve Bayes Classifier yang sudah dijelaskan sebelumnya. Perhitungan tersebut dilakukan dengan menggunakan library Numpy yang disediakan oleh Python.

3. Analisis Faktor

Setelah tahap analisis sentimen selesai, maka dilakukan perhitungan analisis faktor dengan dari seleksi fitur yang berbentuk DTM sebelumnya. Dari seleksi fitur tersebut dilakukan penentuan jumlah faktor dan dilakukan pengelompokkan variabel.

4. Validasi Pakar

Hasil klasifikasi sentimen dan analisis faktor akan di validasi oleh pakar.

HASIL DAN PEMBAHASAN

Adapun beberapa penelitian terdahulu yang memiliki hubungan dengan analisis sentimen pada media sosial. Analisis sentimen dilakukan terhadap kualitas tempat wisata, data yang digunakan berasal dari API Google Maps yang terdapat *convert emoticon emoticon* pada tahap *preprocessing* [8]. Penelitian ini menghasilkan persentase sentimen positif dan negatif sama dengan penelitian sebelumnya [1] dan metode NBC berhasil mengklasifikasikan suatu opini berupa kelas positif dan negatif secara akurat.

Penerapan penelitian text mining juga diaplikasikan pada review film menggunakan algoritma Support Vector Machine [9], dimana data diambil dari komentar masyarakat di media sosial Twitter tentang film populer pada tahun 2016. Data tersebut berupa file CSV yang diambil melalui Twitter API dengan proses crawling. Dataset berjumlah 3000 tweets. Klasifikasi sentimen dibagi menjadi tiga, yaitu opini positif, negatif dan netral. Hasil dari penelitian ini adalah klasifikasi

berdasarkan kata kunci yang dimasukkan user dan pengujian dengan algoritma Support Vector Machine lebih akurat menggunakan jumlah data latih yang banyak.

Metode K-Means Clustering juga dilakukan dalam analisis sentimen untuk automasi penentuan tren topik skripsi, penelitian ini dihasilkan bahwa algoritma K-Means Clustering terbukti dapat membantu proses pengelompokkan berbagai topik skripsi yang ada sehingga diperoleh informasi yang sesuai untuk tren penelitian Universitas dari tahun ke tahun [10]. Penelitian analisis sentimen juga diaplikasikan ke dalam jasa ojek online menggunakan klasifikasi Naïve Bayes Classifier [11]. Data yang dikumpulkan adalah sebanyak 3000 data bersumber dari mention masyarakat tentang gojek dan grab dari akun @gojekindonesia dan @GrabID. Pada penelitian ini dihasilkan persentase akurasi pada klasifikasi.

Penelitian analisis sentimen terhadap suatu produk juga dilakukan menggunakan klasifikasi algoritma Naïve Bayes Classifier [12]. Pengumpulan data dilakukan dengan menarik data tweet dari server Twitter menggunakan API. Data tweet tersebut disimpan kedalam MongoDB. Dihasilkan tiga kategori pada data Twitter, yaitu kategori positif, negatif dan netral.

Penelitian serupa, yaitu analisis sentimen terhadap pemerintahan Jokowi Pemerintahan tersebut dalam bidang politik, ekonomi dan hukum. Hasilnya hanya sentimen positif dan sentimen negatif maka untuk bidang politik dan ekonomi sentimen positif lebih tinggi, sedangkan untuk bidang hukum sentimen negatif lebih tinggi [13]. Media sosial instagram juga diteliti dengan kasus cyberbullying, untuk mengetahui setiap sentimen komentar tersebut, penelitian ini menggunakan fitur Term Frequency-Inverse Document Frequency dan metode klasifikasi Support Vector Machine. Hasil yang didapatkan adalah klasifikasi berupa kelas positif dan negatif dari

cyberbullying, serta menampilkan akurasi yang baik [14].

Telah dihasilkan sistem implementasi *text mining* untuk analisis opini publik terhadap elektabilitas calon presiden 2019 melalui media sosial Twitter yang memberikan hasil yang cukup baik, pengujian hasil dilakukan dengan metode *white box* dan *black box*. Metode *white box* dilakukan dengan menguji apakah sistem pada program berjalan sesuai dengan perhitungan algoritma Naïve Bayes Classifier yang sudah dijelaskan sebelumnya atau tidak dan dengan dilakukan perhitungan akurasi setiap topik dengan persamaan (6).

$$Akurasi\ Topik\ n = \frac{\sum nn + \sum pn}{cn} \quad (6)$$

Dimana

Akurasi Topik n : Akurasi sentimen setiap topik

$\sum nn$: Jumlah klasifikasi benar tweet data uji negatif

$\sum pn$: Jumlah klasifikasi benar tweet data uji positif

Cn : Jumlah keseluruhan tweet data uji Nilai akurasi setiap topik didapatkan

Jika akurasi setiap sentimen per topik sudah dihitung, maka dilakukan perhitungan untuk akurasi keseluruhan topik dengan persamaan (7).

$$Rerata\ Akurasi = \frac{\sum_{n=1}^2 Akurasi\ topik\ n}{n} \quad (7)$$

Hasil Analisis Sentimen

Hasil *preprocessing* dapat menghilangkan format-format karakter yang tidak diinginkan, menghilangkan stopword dan mengubah beberapa kata yang memiliki imbuhan menjadi kata dasar. Namun hasil *preprocessing* ini memiliki kelemahan atau kurang sempurna karena tidak bisa mengontrol singkatan, bahasa lokal, lebay dan tata bahasa yang tidak benar. Selain itu kekurangan dari proses *preprocessing* ini adalah waktu eksekusi yang kurang

cepat, dimana untuk setiap tweet dibutuhkan waktu kurang lebih 0.37 detik hingga proses *preprocessing* selesai. Hal ini mengakibatkan eksekusi untuk keseluruhan data latih positif 'Jokowi' berjumlah 3.642 memakan waktu 15,7 menit, negatif 'Jokowi' berjumlah 2.885 memakan waktu 12,86 menit, positif 'Prabowo' sebanyak 1.836 memakan waktu 17,2 menit dan negatif 'Prabowo' berjumlah 2.214 memakan waktu 11,65 menit.

Hasil Klasifikasi

Hasil klasifikasi dikatakan baik untuk mengklasifikasikan data uji, dan diuji dengan perhitungan akurasi. Angka akurasi menunjukkan probabilitas keberhasilan sistem dalam mengklasifikasikan tweet yang dihitung dari perbandingan jumlah klasifikasi yang benar terhadap seluruh data sampel. Akurasi topik 'jokowi' dapat dihitung sebagai berikut:

$$Akurasi\ topik\ 'jokowi' = \frac{1076}{1214} \times 100\% = 89,70\%$$

Sedangkan pada topik 'prabowo', akurasi dihitung sebagai berikut:

$$Akurasi\ topik\ 'prabowo' = \frac{543}{613} \times 100\% = 84,09\%$$

Dari kedua data tersebut, keseluruhan akurasi dapat dihitung sebagai berikut:

$$Rerata\ akurasi = \frac{89.70 + 84.09}{2} \times 100\% = 86,89\%$$

Tingkat akurasi klasifikasi total dari penelitian yang terdiri sentimen Jokowi dan Prabowo positif dan negatif adalah 86.89%. Artinya, 86.89% dari hasil klasifikasi sesuai dengan karakteristik sentimen. Akurasi ini bergantung pada algoritma Naïve Bayes Classifier dan kualitas data latih yang

ada. Semakin bertambah data latih maka tweet yang terkumpul semakin variatif dan banyak istilah-istilah yang dapat mempengaruhi klasifikasi besar pula akurasi yang didapatkan. Hal yang perlu diperhatikan pula adalah keberadaan sentimen netral. Sentimen netral diartikan sebagai data yang tidak memiliki sentimen. Sebagian besar sentimen netral berisi seperti hal-hal yang berkaitan dengan promosi, pemberitahuan ataupun kegiatan. Dalam penerapannya pula, sentimen netral sangat mengganggu karena tidak dibutuhkan dalam klasifikasi sentimen positif dan negatif.

Hasil Analisis Faktor

Hasil DTM digunakan untuk analisis faktor. DTM memiliki frekuensi kemunculan yang banyak pada tweet yang berbeda-beda. Selanjutnya, kata yang memiliki frekuensi dalam bentuk matrix dilakukan perhitungan analisis faktor menghasilkan nilai komponen (PC) yang menjadi acuan dalam penentuan pengelompokan suatu faktor setiap tweet yang berada sesuai dengan nilai PC yang paling tinggi.

Pada penelitian ini masing-masing sentimen per topik dihasilkan 20 faktor yang sudah diinterpretasikan oleh pakar. Sentimen Jokowi negatif yakni tentang politik, sifat etika, perkataan yang kasar, penghinaan pendidikan, penyebaran pidato hoax, permusuhan pasangan calon, agama, politikus yang munafik, perasaan, bahaya provokasi, hoax, kewaspadaan strategi, rakyat jelata, sifat cebongers, teroris santri, kriminalisasi, kondisi perumahan, fitnah dalam demo, dana hibah desa dan kekalahan dan kampanye. Sentimen Jokowi positif yakni tentang prestasi ekonomi, kualitas maritime, etika berdemokrasi, elektabilitas partai, kebijaksanaan, sifat merangkul semua, rumah yang sempurna, berbahasa baik dan benar, optimis untuk menang, kemenangan survey, pemilihan Jokowi, kemenangan partai, infrastruktur, kompetisi turnamen, perevolusian,

apresiasi impor, swasembada pangan, pembangunan desa, birokrasi debat Jokowi dan program proyek Jokowi.

Sedangkan sentimen Prabowo negatif yakni tentang keagamaan, emosional, demo joget dalam debat, penghinaan kemenkeu, swasembada tanah, halusinasi argument, kalimat pembodohan, pembulian terhadap para kece bong, berita hoax, menyakiti kemenkeu, ketidakadilan oposisi, kekuasaan yang konyol, fitnah tentang revolusi, perbohongan koalisi, kemunafikan, perkataan kasar, korupsi partai, sandiwara pribadi, persepsi argument dan etika debat. Sentimen Prabowo positif yakni tentang pembicaraan baik, sosialisasi anak muda, kerukunan, berdoa, kampanye, pribadi yang ganteng dan mandiri, penghormatan, konsisten dalam komitmen, melantik tentara, berprestasi, kabinet bangga kepada pilpres, sejarah pertahanan yang bagus, emak-emak bagus, elektabilitas partai, komando pilot yang tulus, semangat mengkritik lawan, martabat yang mantap, jenderal yang menguasai lapangan, debat bangsa, sujud dan syukur selalu.

SIMPULAN

Berdasarkan hasil penelitian dan uji coba yang dilakukan terhadap sistem analisis opini publik terhadap elektabilitas calon Presiden 2019, dapat ditarik kesimpulan sebagai berikut.

1. Sistem berhasil mengumpulkan tweet secara realtime.
2. Teknik preprocessing untuk membersihkan tweet asli dilakukan dengan baik.
3. Seleksi fitur berhasil dilakukan, yakni menghapus jumlah kata yang kurang dari tiga.
4. Metode Naïve Bayes Classifier berhasil dilakukan dengan menghitung nilai \ln probabilitas untuk mengklasifikasikan tweet yang bersentimen dengan hasil uji akurasi 89.70% untuk topik Jokowi, 84.09% untuk topik Prabowo pada

1.352 tweet dan rerata akurasi 86.89% pada 3.528 tweet

SARAN

Sistem berhasil mengelompokkan faktor berdasarkan nilai PC yang paling besar dengan teknik analisis faktor yang mengelompokkan kata-kata dalam dimensi yang lebih kecil dan diperlukan masukan jumlah faktor oleh user agar dapat memberikan hasil yang maksimal

DAFTAR PUSTAKA

- [1] Pramono, A., Indriati, R. & Nugroho, A., 2017. Sentiment Analysis Tokoh Politik Pada Twitter. Seminar Nasional Inovasi Teknologi UN PGRI Kediri, Volume 1 No 1, pp. 189-194.
- [2] Nurhuda, F., Sihwi, S. W. & Doewes, A., 2013. Analisis Sentimen Masyarakat Terhadap Calon Presiden Indonesia 2014 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier. ITSMART: Jurnal Teknologi dan Informasi, pp. 35-42.
- [3] Koncz, P. & Paralic, J., 2011. An approach to feature selection for sentiment analysis. Poprad, Slovakia: IEEE International Conference on Intelligent Engineering Systems.
- [4] Indriani, A., 2014. Klasifikasi Data Forum dengan menggunakan Metode Naive Bayes Classifier. Jurnal Fakultas Hukum UII, pp. 6-6.
- [5] Pradhan, V. M., Vala, J. & Balani, P., 2016. A Survey on Sentiment Analysis Algorithms for Opinion Mining. International Journal of Computer Applications, pp. 133(9):7- 11.
- [6] Amerika, S. & Nurhidayah, 2014. Analisis Faktor-Faktor yang Mempengaruhi Kinerja Kemitraan Bisnis PT PLN. Jurnal Manajemen dan Akuntansi, p. 3(2).
- [7] Ary, M., 2016. Analisis Faktor Pemilihan Program Studi Untuk Meraih Keunggulan Bersaing. Jurnal Informatika, 3(1), pp. 83-84
- [8] Wilianto, L., Pudjiantoro, T. H. & Umbara, F. R., 2017. Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Naive Bayes Classifier Studi Kasus Jawa Barat. Prosiding SNATIF, pp. 439-448.
- [9] Rahutomo, F., Saputra, P. Y. & Fidyawan, M. A., 2018. Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine. Jurnal Informatika Polinema, 4(2), pp. 93-93.
- [10] Prilianti, K. R. & Wijaya, H., 2014. Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering. Jurnal Cybermatika, p. 2(1).
- [11] Nugroho, G., Chrisnanto, Y. H. & Wahana, A., 2016. Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode Naive Bayes. Prosiding SNST Fakultas Teknik, 1(1).
- [12] Retnawiyati, E., M.Kom, F. M. & Negara, M.Kom, E. S., 2016. Analisis Sentimen Pada Data Twitter dengan Menggunakan Text Mining terhadap Suatu Produk.
- [13] Budi, I. & Yislam, 2015. Analisis Sentimen Masyarakat Terhadap Pemerintahan Jokowi Menggunakan Data Twitter.
- [14] Luqyana, W. A., Cholissodin, I. & Perdana, R. S., 2018. Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer 2(11), pp. 4704-4713.