

Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa Barat 2018 di Twitter Menggunakan Naive Bayes

Haga Simada Ginting¹, Kemas Muslim Lhaksana², Danang Triantoro Murdiansyah³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

⁴Divisi Digital Service PT Telekomunikasi Indonesia

¹hagasimada@gmail.com, ²kemas.muslim@gmail.com, ³dto.lecture@gmail.com

Abstrak

Pemilihan kepala daerah (pilkada) merupakan pemilihan umum untuk memilih gubernur dan wakil gubernur yang dilakukan oleh masyarakat setempat yang memenuhi syarat sebagai pemilih. Gubernur merupakan pemimpin daerah yang bertugas dalam memimpin suatu wilayah daerah provinsi di Indonesia. Dalam hal ini bakal calon gubernur membutuhkan sentimen dari masyarakat sebagai sumber informasi untuk mengetahui citra bakal calon gubernur. Sentimen yang didapatkan dari masyarakat tidak hanya bersifat positif, melainkan juga bersifat negatif dan netral. Sentimen yang digunakan pada penelitian ini adalah tweet dari masyarakat yang dicrawling dari twitter dan berhubungan dengan bakal calon gubernur Jawa Barat 2018. Pada penelitian tugas akhir ini penulis membangun sistem klasifikasi sentimen masyarakat dengan metode Naive Bayes Classifier (NBC). Model Naive Bayes Classifier (NBC) digunakan untuk mendapatkan nilai preference value dari masyarakat terhadap kandidat calon gubernur Jawa Barat 2018. Hasil pengujian dengan metode evaluasi menghasilkan rata-rata akurasi sebesar 76,56%. Untuk hasil pengujian respon positif masyarakat di twitter yang terbesar diperoleh oleh Deddy Mizwar dengan nilai preference value sebesar 34,3%. Dengan demikian, klasifikasi sentimen menggunakan NBC dapat digunakan untuk mengukur preference value pada kasus pemilihan kepala daerah.

Kata Kunci: Twitter, Pre-processing, Naive Bayes Classifier, Cross validation, Analyzing Sentiment, Preference Value

Abstract

Local elections are the general election for choose governor and vice governor that held by society who fulfill the requirement as a voter. Governor is a leader of province region in Indonesia. In this case the candidates need sentiment from society as the source of information for knowing image of the governor candidates. The sentiments that has got from society are not only positive, it can also negative and neutral. Sentiments that used for this research are tweet from the society that the result of crawling from twitter that related with the governor candidates of west java 2018. In this research, author build the society sentiment classification system with Naive Bayes Classifier. Naive Bayes Classifier used for get preference value from society to the governor candidate of west java 2018. The testing result with evaluation method got 76,56% accuracy. The highest testing result for society's positive response on twitter goes to Deddy Mizwar with 34,4% preference value. So, the sentiment classification with NBC can be used for measure preference value in local election case.

Keywords: Twitter, Pre-processing, Naive Bayes Classifier, Cross validation, Analyzing Sentiment, Preference Value

1. Pendahuluan

Latar Belakang

Pemilihan kepala daerah (pilkada) merupakan pemilihan umum untuk memilih kepala daerah dan wakil kepala daerah wilayah tertentu yang dilakukan oleh masyarakat setempat yang memenuhi syarat sebagai pemilih. Gubernur merupakan pemimpin daerah yang bertugas dalam memimpin suatu wilayah daerah provinsi di Indonesia. Salah satu fungsi tugas dari gubernur adalah menyebarluaskan informasi, mengkomunikasikan kebijakan, rencana kerja, dan capaian kinerja kepada masyarakat luas, melalui media tradisional, media konvensional, dan media internet.

Komunikasi dengan menggunakan media internet dapat menjangkau luas dan cepat kepada semua pihak yang saat ini telah dapat dinikmati secara luas. Masyarakat telah mengalami perubahan pola pikir ke arah yang semakin kritis dalam menanggapi kondisi yang ada, dimana dengan adanya kondisi seperti ini maka kinerja kepala daerah dituntut untuk mampu memenuhi berbagai kebutuhan masyarakat dalam segala aspek. Media sosial merupakan salah satu media yang digunakan masyarakat untuk mengkritik kinerja pemerintahan. Media sosial adalah sebuah aplikasi berbasis internet dengan dasar ideologi dan teknologi Web 2.0 dan memungkinkan penciptaan dan pertukaran user-generated content. Media sosial kini marak digunakan karena kemudahan dalam mengakses dan minimnya biaya yang dibutuhkan. Kemudahan akses yang dimiliki bukan hanya karena bisa digunakan di mana saja, namun juga perangkat yang kompatibel tersedia sedemikian banyak. Kemampuan dari media sosial yang bisa menghasilkan komunikasi dua arah adalah salah satu kelebihanannya. Selain itu, sifat media sosial yang mampu membangun komunikasi dua arah memungkinkan media sosial menjadi wadah untuk berbagi

ide dan informasi, bekerja sama, menemui orang baru, berkomunikasi, sehingga membuatnya hampir menyerupai interaksi manusia biasa.

Media sosial yang dimaksud pada penelitian kali ini adalah twitter. Twitter adalah sebuah layanan dari jaringan internet dan mikroblogging yang memiliki fitur yang memungkinkan penggunaannya untuk membuat dan membaca status. Status yang dapat ditulis oleh pengguna maksimum sebanyak 140 karakter, nama dari status ini biasa disebut kicauan (*tweet*) [3]. Tweet dapat dilihat secara publik, namun pengguna dapat membatasi pengiriman status ke daftar teman yang diinginkan. Pengguna juga dapat melihat kicauan orang lain dengan menjadi pengikut akun orang yang ingin dilihat kicauannya. Beberapa simbol yang sering dipakai di twitter adalah '@' simbol ini biasanya digunakan untuk menyebutkan atau membalas kicauan pengguna lain, dan '#' biasa disebut hastag berfungsi sebagai topik pembahasan yang akan di bahas [3].

Sehingga pada penelitian ini akan dilakukan analisis sentimen dengan menganalisis tweet-tweet yang memension kandidat bakal calon gubernur Jawa Barat tahun 2018. Analisis sentimen merupakan sebuah cabang penelitian di domain text mining yang mulai terkenal pada awal tahun 2002-an. Analisis sentiment sendiri sebenarnya merupakan proses untuk mengidentifikasi dan mengekstrak data sentimen yang biasanya akan dikategorikan berdasarkan polaritasnya, apakah itu positif, negatif, dan netral. Ketika kita mendekati pemilihan gubernur, sentimen terhadap kandidat akan mempengaruhi pemenang dalam pemilihan gubernur. Penulis tertarik pada bagaimana masyarakat memandang kandidat bakal calon gubernur Jawa Barat tahun 2018. Adapun nama kandidat calon gubernurnya adalah Deddy Mizwar, Dedi Mulyadi, Ridwan Kamil.

Data-data dari opini masyarakat yang di twitter bisa kita olah dengan teknik data mining yaitu klasifikasi. Untuk pengelompokkan sentimen masyarakat penulis membagi tiga indikator yaitu sentimen positif, netral, dan negatif, dengan indikator berdasarkan dari tweet yang dilontarkan masyarakat yang berhubungan dengan kandidat. Klasifikasi yang akan dilakukan dalam penelitian ini yaitu klasifikasi sentimen masyarakat menggunakan metode Naive Bayes Classifier. Penelitian terkait sebelumnya pernah dilakukan oleh Alec Go dan kawan-kawan [1]. Setelah mendapatkan sistem classifier, maka akan dilanjutkan dengan pencarian preference value. Preference value dilakukan untuk mengetahui trending atau respon positif dari masyarakat terhadap bakal calon gubernur Jawa Barat 2018 di twitter.

Topik dan Batasannya

Berdasarkan latar belakang yang telah dijelaskan di awal maka didapat beberapa rumusan masalah, yaitu bagaimana mendapatkan model klasifikasi sentimen menggunakan Naive Bayes Classifier pada twitter dan bagaimana preference value masyarakat terhadap bakal calon gubernur Jawa Barat tahun 2018 di twitter?

Batasan masalah pada penelitian ini adalah data yang digunakan diambil hanya dari twitter. Data yang digunakan berupa tweet, user, dan retweet yang diambil dari tanggal 29 Agustus 2017 sampai dengan 14 Desember 2017. Bakal calon gubernur Jawa Barat 2018 yang akan digunakan dalam penelitian ini yaitu: Ridwan Kamil, Deddy Mizwar, dan Dedi Mulyadi berdasarkan info dari berbagai sumber mainstream pada saat penelitian ini dikerjakan (sekitar agustus 2017).

Tujuan

Dengan adanya masalah yang telah disebutkan, maka tujuan dari penelitian ini yaitu mendapatkan model klasifikasi menggunakan Naive Bayes Classifier pada twitter. Mendapatkan nilai preference value masyarakat terhadap bakal calon gubernur Jawa Barat tahun 2018 di twitter.

2. Studi Terkait

2.1 Naive Bayes Classifier

Naive Bayes Classifier (NBC) adalah metode yang digunakan untuk mencari nilai probabilitas tertinggi dalam proses klasifikasi data uji pada kategori (kelas) yang paling tepat. Dalam penelitian tugas akhir ini, yang menjadi data uji maupun data latih adalah data user dengan fitur-fiturnya. Ada dua tahap dalam NBC untuk proses klasifikasinya, yaitu training dan testing [12]. Tahap pertama adalah pelatihan terhadap *sentiment* yang sudah diketahui kelasnya untuk membangun sebuah model probabilistik. Sedangkan tahap kedua adalah proses klasifikasi *sentiment* yang belum diketahui kebenaran dari kelasnya [2][12][13].

Secara matematis, NBC dituliskan sebagai berikut:

$$P(A|B) = \frac{(P(B|A) \times P(A))}{(P(B))} \quad (1)$$

Keterangan:

- A = sampel data yang label kelasnya tidak diketahui.
- B = kelas-kelas hasil klasifikasi.

$P(A|B)$ = probabilitas terjadinya A jika B diketahui. Disebut probabilitas posterior, karena peluang A bergantung dari nilai B tertentu.

$P(B|A)$ = probabilitas terjadinya B jika A diketahui, disebut likelihood function, karena peluang B tergantung dengan peluang data *sample* A.

$P(A)$ = probabilitas A merupakan probabilitas dari sample yang mempunyai kelas A.

$P(B)$ = probabilitas prior B, dan bertindak sebagai normalizing constant. Secara intuitif, teorema Bayes menggambarkan bahwa perubahan pada "A" dapat diamati apabila "B" terlebih dahulu diamati.

2.2 Measuring Performance

Performance Measure adalah tahap analisis dan evaluasi terhadap performansi sistem yang kita rancang. Dalam penelitian tugas akhir ini, digunakan performansi diukur dari precision, recall, dan nilai akurasi..

Precision adalah jumlah user yang dengan benar diklasifikasikan dalam sebuah kelas dibagi dengan jumlah total user yang diklasifikasikan dalam kelas tersebut [9]. Precision juga sering disebut sebagai tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Berikut rumus dari precision:

$$precision = \frac{TP}{(TP + FP)} * 100\% \quad (2)$$

Recall adalah jumlah user yang dengan benar diklasifikasikan dalam sebuah kelas dibagi dengan jumlah total user dalam kelas tersebut [9]. Recall juga sering disebut sebagai tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Berikut rumus dari recall:

$$Recall = \frac{TP}{(TP + FN)} * 100\% \quad (3)$$

Akurasi merupakan parameter evaluasi terhadap sistem yang dibangun dalam penelitian tugas akhir ini. Akurasi sering disebut sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Berikut rumus dari akurasi[12] :

$$akurasi = \frac{jumlah\ kategorisasi\ benar}{jumlah\ dokumen\ uji} * 100\% \quad (4)$$

2.3 Prefrence Value

Prefrence value adalah penilaian untuk mengetahui jumlah respon positif pengguna twitter. Dalam tugas akhir ini nilai Prefrence value digunakan untuk mengetahui trending bakal calon gubernur Jawa Barat 2018 di twitter. Semakin tinggi nilai Prefrence value yang dihasilkan maka dapat diartikan semakin tinggi pula respon positif yang didapatkan[16]. Berikut rumus dari preference value:

$$Prefrence\ value = \frac{sp}{sp + sr + sn} * akurasi \quad (5)$$

Keterangan:

Sp:sentimen positif

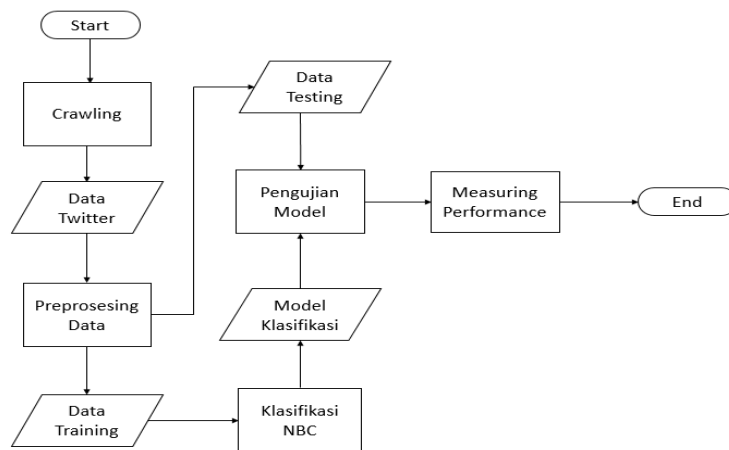
Sn:sentimen negatif

Sr:sentimen netral

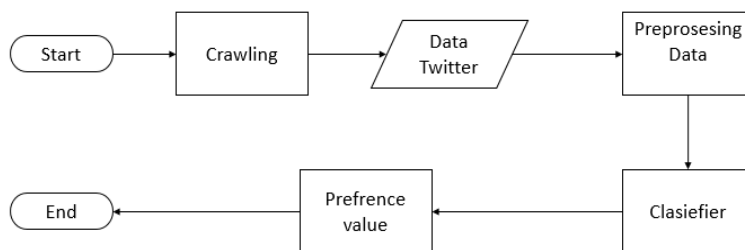
3. Sistem yang Dibangun

Penelitian ini dilakukan untuk mengklasifikasikan sentimen-sentimen yang terdapat pada tweet yang diposting oleh masyarakat di twitter yang berhubungan dengan bakal calon gubernur Jawa Barat 2018 berdasarkan. Dimana dari data status yang didapat akan diklasifikasikan menjadi beberapa jenis kategori sentimen yaitu positif, negatif dan netral. Metode yang digunakan untuk mengklasifikasi adalah metode Naive Bayes Classifier. Sebelum melakukan klasifikasi diperlukan preprocessing data untuk menyederhanakan data.

Setelah mendapatkan sistem clasifier yang sudah diuji pada cross validation, dilanjutkan dengan mencari preference value. Rangkaian proses dalam sistem yang akan dibangun seperti pada gambar 1 dan gambar 2.



Gambar 1 Flowchart Sistem Klasifikasi



Gambar 2 Flowchart Preference Value

3.1 Crawling

Crawling Data merupakan suatu teknik pengambilan data secara otomatis dari suatu tempat penyimpanan data. Pada penelitian ini objek dari tahapan crawling data adalah laman web milik Twitter. Data yang didapat dari tahapan ini berupa user dan tweet. Selanjutnya dengan menggunakan API Twitter yang sudah diimplementasikan dalam bahasa pemrograman PHP, data tweet akan didapatkan dengan cara mengunduh secara otomatis dari data user yang didapat [10]. Adapun data yang akan di crawling adalah yang memension tweet dari bakal calon gubernur Jawa Barat tahun 2018. Keluaran dalam tahap ini adalah data user.

3.2 Pre-processing

Preprocessing yang dilakukan pada penelitian kali ini pertama dilakukan *cleansing* yang bertujuan untuk membersihkan data tweet yang tidak mendukung penelitian ini [15]. Setelah cleansing dilakukan teknik *case folding* yang bertujuan untuk merubah huruf besar menjadi huruf kecil[5]. Setelah *case folding* dilakukan teknik *filtering* yang bertujuan untuk membuang kata-kata yang kurang penting. Tahap terakhir adalah tokenisasi yang bertujuan untuk memecah teks menjadi kata-kata yang berurutan[7].

3.3 Data training dan testing

Data training dan testing yang akan digunakan adalah data yang sudah di pre-processing. Data berupa tweet dari pengguna twitter. Pemilihan data yang akan digunakan berdasarkan pada tweet dan retweet yang berhubungan dengan kandidat. Dimana data training berfungsi untuk membentuk model yang akan digunakan dan data testing berfungsi untuk mengetahui performansi dari model yang telah terbentuk.

3.4 Klasifikasi NBC

Naive Bayes Clasifier pada penelitian kali ini bertujuan untuk menghitung probabilitas bersyarat pada tiap atribut(kata) tiap kelas. Tahapan- tahapan yang dilakukan pada proses klasifikasi adalah pertama *bag of words* bertujuan untuk mengumpulkan kata yang ada pada setiap *tweet* berdasarkan frekwensi kemunculan di *tweet* tersebut. Setelah *bag of words* tahapan selanjutnya adalah training hal ini dilakukan untuk mendapatkan model dalam bentuk probabilitas. Untuk menghindari peluang 0 yang ada ditraining dan merusak hasil testing sehingga dibutuhkan *laplace smoothing*. *Laplace smoothing* adalah menambahkan angka 1 dibagi jumlah semua fitur yang ditambahkan ke semua fitur sehingga tidak ada yang nilainya 0 [11]. Setelah mendapatkan model probabilitas dari training tahapan selanjutnya testing bertujuan untuk menguji hasil dari training dengan menggunakan data yang berbeda.

3.5 Measuring Performance

Data yang sudah dikategorikan positif, netral, dan negatif oleh NBC akan diuji tingkat keakurasiannya beserta nilai recall dan precision-nya dengan menggunakan metode Cross Validation dengan sepuluh folds. [9]. *10-fold validation* adalah membagi data menjadi 10 bagian setiap bagian memiliki jumlah data yang sama dimana pengujiannya dilakukan 9 bagian menjadi data training dan 1 bagian menjadi data testing proses pengujian akan diulang sampai 10 kali.

3.6 Preference Value

Setiap kandidat akan dihitung nilai preference value-nya melalui tweet yang dikumpulkan yang berbeda dengan data training. Dimana tweet yang digunakan di crawling dari tanggal 7 Desember 2017 sampai dengan 14 Desember 2017. Tweet yang didapat akan dimasukkan ke model klasifikasi. Setelah diklasifikasikan maka akan dilakukan perhitungan preference value-nya. Hal ini dilakukan untuk mengetahui trending bakal calon gubernur Jawa Barat 2018 di twitter.

4. Evaluasi

4.1 Data Acuan

Data acuan yang digunakan pada model yang akan dibangun berupa data tweet dan retweet yang diambil pada tanggal 29 Agustus 2017-14 Desember 2017. Data tweet yang digunakan untuk membuat model klasifikasi ini sebanyak 900 data tweet. Dimana tweet yang digunakan sudah diberi kelas dengan teknik labeling secara manual. Rincian data tweet terdiri dari 300 tweet untuk masing-masing kandidat. Dapat dijabarkan dari 300 tweet ini terdiri dari 100 tweet positif, 100 tweet netral, dan 100 tweet negatif.

Selain untuk membangun model data acuan juga digunakan untuk preference value. Data ini digunakan untuk memprediksi kecondongan politik masyarakat pengguna twitter terutama yang berhubungan dengan bakal calon gubernur Jawa Barat 2018. Adapun jumlah tweet yang digunakan sebanyak tweet, dapat dijabarkan dari 1031 tweet ini terdiri dari 337 tweet mengenai Ridwan Kamil, 335 tweet mengenai Deddy Mizwar, dan 359 tweet mengenai Dedi Mulyadi. Data tweet tersebut diambil secara manual dari alamat situs www.twitter.com.

Data tweet berupa teks tweet berbahasa Indonesia Indonesia beserta fitur tweet (URL, mention, retweet dan tweet biasa) dengan jumlah 12.579 tweet. Data tweet tersebut diambil dari tweet pengguna twitter yang mengeluarkan pendapat tentang kandidat calon gubernur Jawa Barat tahun 2018. Data tweet yang diambil menggunakan API twitter yang sudah diimplementasikan dalam bahasa pemrograman PHP. Data tweet disimpan dalam korpus tweet. Berikutnya data tweet yang di dapat akan disimpan dalam database menggunakan MySQL.

4.2 Implementasi Sistem

Berikut adalah implementasi dari rancangan model sistem yang digunakan dalam bahasa pemrograman PHP dengan perangkat lunak visual code, serta basis data menggunakan XAMPP dan MySQL server. Pada gambar 3 adalah user interface dari proses crawling data.



Gambar 3 User Interface Crawling Data

Data twitter yang didapatkan dari proses crawling akan disimpan didalam database. Data yang disimpan dalam database berupa *id_tweet*, *id_akun*, *akun*, *tweet*, dan *created* (tanggal postingan tweet tersebut). Gambar 4 merupakan contoh dari *user interface* yang digunakan untuk klasifikasi sentimen. Pada gambar 4 adalah user interface Ridwan Kamil untuk user interface Deddy Mizwar dan Dedi Mulyadi sama halnya dengan user interface Ridwan Kamil.



Gambar 4 User interface ridwan Kamil

User interface ini berfungsi untuk mengklasifikasikan data tweet yang berhubungan dengan Ridwan Kamil. Cara kerja sistem klasifikasi ini berdasarkan dari probabilitas kata-kata yang terdapat dari tweet tersebut. Probabilitas tertinggi dari ketiga kategori (positif, negatif, netral) ini akan menentukan sentimen pada tweet tersebut. Untuk kasus pada gambar 4 *tweet* tersebut berisi sentimen positif dengan probabilitas 89,25 %.

4.3 Pengujian Sistem

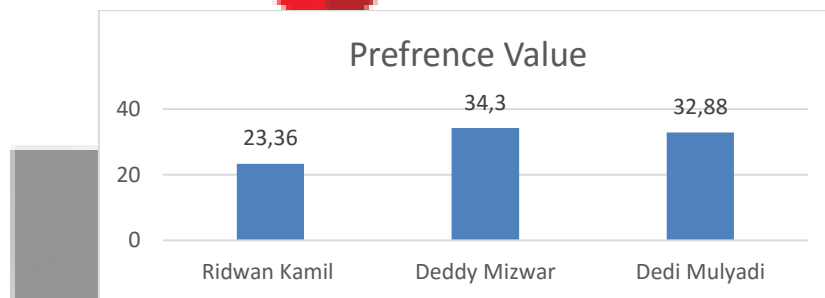
Pada penelitian ini menggunakan persamaan (4) untuk menghitung akurasi. Akurasi bertujuan untuk mengetahui performansi parameter akurasi dari klasifier yang dibangun. Metode evaluasi yang digunakan dalam pengujian ini adalah 10-fold validation, dimana metode ini membagi data menjadi 10 bagian dengan rincian 9 bagian untuk data training dan 1 bagian untuk data testing. Lalu diuji sebanyak 10 kali dengan kombinasi data training dan data testing yang berbeda. Akurasi yang didapatkan merupakan rata-rata dari 10 akurasi pengujian. Sehingga didapatkan akurasi sistem untuk clasifier Ridwan Kamil dari sepuluh skenario yang dilakukan adalah 80,33%. Untuk rata-rata akurasi clasifier Deddy Mizwar didapat sebesar 73,67%, dan untuk rata-rata akurasi clasifier Dedi Mulyadi sebesar 75,67%. Sehingga dapat dikalkulasikan rata-rata akurasi sistem keseluruhan adalah 76,56%. Selain mencari nilai akurasi metode pengujian yang untuk mengetahui performansi dari sistem yang dibangun adalah *recall* dan *precision*. Untuk hasil lengkap dari pengujian sistem keseluruhan dapat dilihat pada tabel 1 *Measuring Performance*.

Tabel 1 *Measuring Performance*

Kandidat	Recall			Precision			Akurasi
	Positif	Netral	Negatif	Positif	Netral	Negatif	
Dedi Mulyadi	79,69%	76,34%	78,59%	87%	80%	60%	75.67%
Ridwan Kamil	80,8%	79,31%	87,33%	89%	94%	58%	80.33%
Deddy Mizwar	75,95%	73,38%	85,71%	81%	48%	92%	73.67%

4.4 Preference Value

Tujuan dari preference value untuk mengetahui respon positif dari pengguna twitter, sehingga dapat dilihat respon positif bakal calon gubernur Jawa Barat 2018. Berikut adalah grafik perbandingan preference value untuk bakal calon gubernur Jawa Barat 2018 Ridwan Kamil, Deddy Mizwar, dan Dedi Mulyadi dalam jangka waktu 7-14 Desember 2017 yang dapat dilihat pada gambar 5.

Gambar 5 Perbandingan *Preference value*

Berdasarkan gambar 4.8 dapat dilihat bahwa Deddy Mizwar merupakan bakal calon gubernur Jawa Barat 2018 yang mendapatkan respon positif paling tinggi dengan nilai *preference value* sebesar 34,3. Untuk respon positif terkecil diperoleh Ridwan kamil dengan nilai *preference value* sebesar 23,36. Sehingga dapat disimpulkan bahwa Deddy Mizwar merupakan bakal calon gubernur Jawa Barat 2018 yang mendapatkan respon positif tertinggi berdasarkan data tweet yang diambil dari tanggal 07-14 Desember 2017.

4.5 Analisis Hasil

Pada penelitian ini performansi sistem akan diukur dengan metode akurasi menggunakan persamaan (4). Metode pengujian yang dilakukan adalah 10-fold validation. 10-fold validation adalah membagi data menjadi sepuluh bagian yaitu sembilan bagian sebagai data training dan satu bagian menjadi data testing. Untuk percobaan ini jumlah data yang digunakan adalah sebanyak 900 data tweet yang diambil dari twitter. Tweet yang digunakan adalah tweet-tweet yang berhubungan dengan bakal calon gubernur Jawa Barat 2018. Kandidat yang dimaksud adalah Dedi Mulyadi, Deddy Mizwar, dan Ridwan Kamil. Data tweet yang didapatkan akan diklasifikasikan berdasarkan sentimen positif, sentimen netral, dan sentimen negatif. Sehingga data tweet yang didapatkan untuk masing-masing kandidat sebanyak 300 tweet, dengan rincian 100 tweet untuk sentimen positif, 100 tweet untuk sentimen netral, dan 100 tweet untuk sentimen negatif. Hal ini dilakukan untuk menguji sistem model klasifikasi yang dibuat. Metode klasifikasi yang digunakan untuk pembuatan sistem yang akan digunakan adalah metode Naive Bayes Classifier.

Pada penelitian ini dilakukan dua kali percobaan. Percobaan dilakukan berdasarkan data tweet yang dipilih menurut sentimen yang terdapat pada data tweet yang sudah dikumpulkan melalui teknik crawling. Sehingga didapatkan akurasi dari percobaan pertama dan percobaan kedua, hasil akurasi yang didapatkan dapat dilihat pada tabel 2.

Tabel 2 Perbandingan Performansi Sistem

Percobaan	Ridwan Kamil	Deddy Mizwar	Dedi Mulyadi	Rata-rata akurasi
1	64,67	73,67	66,67	68,34
2	80,33	73,67	75,67	76,56

Berdasarkan tabel 1 dapat dilihat performansi akurasi yang didapatkan dari percobaan pertama untuk bakal calon gubernur Jawa Barat 2018 untuk Dedi Mulyadi didapat hasil akurasi 64,67%, untuk Deddy Mizwar 73,67%, dan untuk Ridwan Kamil 66,67%. Sehingga dapat dirata-ratakan dari percobaan pertama akurasi yang didapatkan sebesar 68,34%. Untuk percobaan kedua performansi akurasi yang didapatkan untuk bakal calon gubernur Jawa barat 2018 untuk Dedi Mulyadi didapat hasil akurasi 75,67, untuk Deddy Mizwar hasil akurasi yang didapatkan sebesar 73,67%, dan untuk Ridwan Kamil hasil akurasi yang didapatkan sebesar 80,33%. Sehingga didapatkan rata-rata akurasi untuk percobaan kedua sebesar 76,56%. Berdasarkan dua percobaan yang dilakukan, percobaan kedua memiliki akurasi yang lebih besar dari percobaan pertama. Hal ini disebabkan karena data yang diambil berbeda secara random.

Hal-hal yang mempengaruhi akurasi yang didapatkan pada percobaan-percobaan yang dilakukan adalah data tweet yang didapatkan banyaknya kata yang disingkat sehingga sistem sulit untuk mengklasifikasikan sentimen yang tertera pada data tweet. Contoh kata-kata yang dimaksud seperti kata “yang” yang disingkat menjadi “yg” atau “y”, “dengan” yang disingkat menjadi “dgn” atau “dg” dan lain-lain. Hal lainnya yang mempengaruhi akurasi yang didapatkan adalah sulitnya untuk menentukan kalimat ambigu menurut sistem, kalimat ambigu yang dimaksud adalah seperti kalimat “saya lebih senang lihat bapak main sinetron daripada jadi gubernur” kalimat tersebut kalau diklasifikasikan secara manual akan menjadi sentimen negatif tetapi untuk sistem yang dibuat hal tersebut belum tentu jadi sentimen negatif. Selain hal yang dijelaskan diatas penyebab perbedaan akurasi pada percobaan pertama dan percobaan kedua adalah data yang diganti secara random. Data yang diganti secara random adalah data yang salah labeling pada saat pemberian label pada *tweet* sehingga setelah percobaan pertama maka data di ganti secara random setelah di ganti maka didapatkan hasil akurasi yang lebih maksimal. Membutuhkan data yang lebih banyak lagi untuk memberikan kategori kata yang lebih bervariasi. Dengan variasi kata yang lebih banyak maka sistem akan lebih mudah untuk mengklasifikasi sentimen yang terdapat pada tweet yang berhubungan dengan bakal calon gubernur Jawa Barat 2018. Sehingga hasil yang didapatkan akan lebih maksimal.

5. Kesimpulan

Berdasarkan dari hasil pengujian, kesimpulan yang didapatkan sebagai berikut:

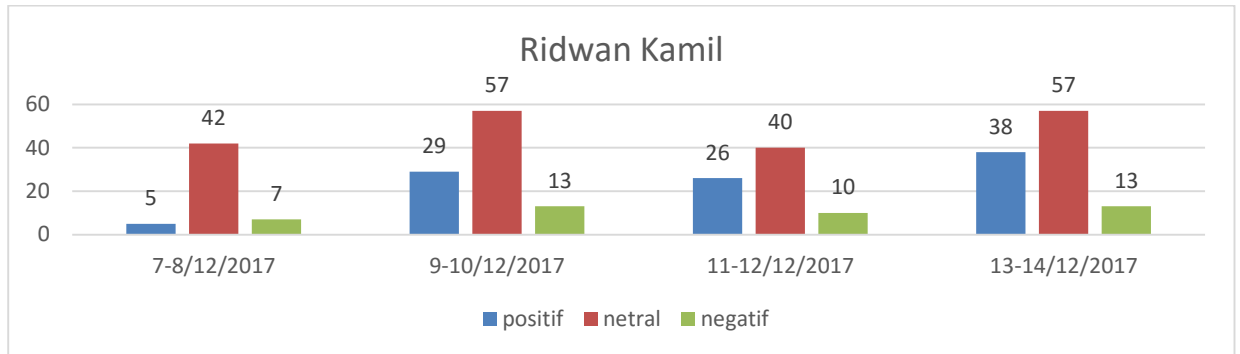
1. Naive Bayes Clasifier dapat digunakan untuk model sistem klasifikasi sentimen pendapat masyarakat di twitter terhadap bakal calon gubernur Jawa Barat 2018 dengan rata-rata akurasi yang didapatkan sebesar 76,56% .
2. Dengan demikian, klasifikasi sentimen menggunakan NBC dapat digunakan untuk mengukur preference value pada kasus pemilihan kepala daerah. Dari hasil preference value bakal calon Deddy Mizwar mendapatkan respon positif sebesar 34,3, untuk Ridwan Kamil sebesar 23,36, dan Dedi Mulyadi sebesar 32,88. Sehingga dapat disimpulkan bahwa Deddy Mizwar mendapatkan respon positif tertinggi, berdasarkan data tweet yang diambil dari tanggal 7 Desember 2017 sampai dengan 14 Desember 2017 .

Daftar Pustaka

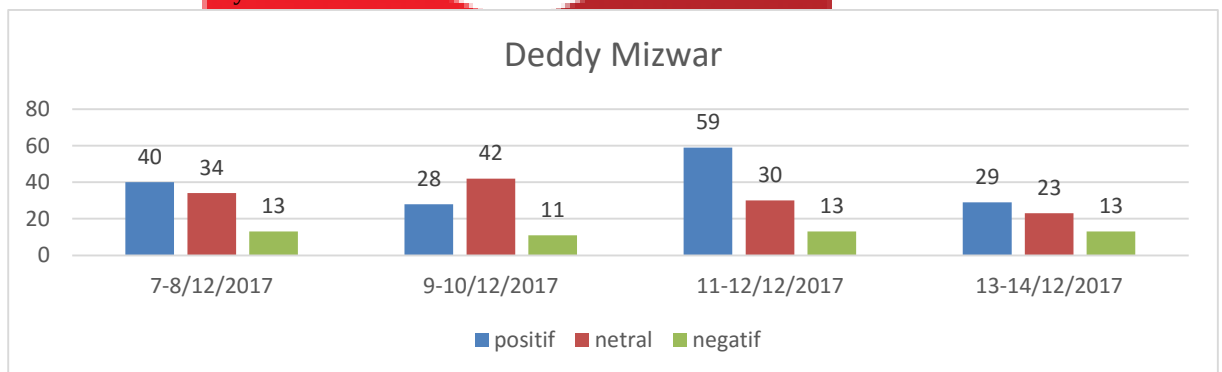
- [1] Criado, J. I., Sandoval-Almazan, R., & Gil-Garcia, J. R. (2013). Government innovation through social media. *Government Information Quarterly*, 30(4), 319-326.
- [2] Sembodo, J. E., Setiawan, E. B., & Baizal, Z. K. A. (2017). A Framework for Classifying Indonesian News Curator in Twitter. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 15(1), 357-364.
- [3] Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. In *Proceedings of the 19th international conference on World wide web* (pp. 591-600). ACM.
- [4] Novantirani, A., Sabariah, M. K., & Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. *eProceedings of Engineering*, 2(1).
- [5] Putranti, N. D., & Winarko, E. (2014). Analisis sentimen twitter untuk teks berbahasa Indonesia dengan maximum entropy dan support vector machine. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 8(1), 91-100.
- [6] Sagayam, R., Srinivasan, S., & Roshni, S. (2012). A survey of text mining: Retrieval, extraction and indexing techniques. *International Journal of Computational Engineering Research*, 2(5).
- [7] Tanesab, F. I. (2017). *Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine* (Doctoral dissertation, Magister Sistem Informasi Program Pascasarjana FTI-UKSW).
- [8] Turney, P. D. (2002, July). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417-424). Association for Computational Linguistics.
- [9] P. Tan, M. Steinbach and V. Kumar, *Introduction to Data Mining*, Boston: Pearson Education, 2006.
- [10] 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).
- [11] Juan, A., & Ney, H. (2002, April). Reversing and Smoothing the Multinomial Naive Bayes Text Classifier. In *PRIS* (pp. 200-212).
- [12] Hamzah, A. (2012, November). Klasifikasi teks dengan naïve bayes classifier (nbc) untuk pengelompokan teks berita dan abstract akademis. In *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III* (No. 2011, pp. 269-277).
- [13] Manurung, R. (2008). Machine learning-based sentiment analysis of automatic indonesian translations of english movie reviews. In *Proceedings of the International Conference on Advanced Computational Intelligence and Its Applications (ICACIA)*, Depok, Indonesia, 2008.
- [14] Chin, D., Zappone, A., & Zhao, J. (2016). Analyzing Twitter Sentiment of the 2016 Presidential Candidates.
- [15] Buntoro, G. A., Adji, T. B., & Purnamasari, A. E. (2014). Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation. *CITEE 2014*, 39-43.
- [16] Endany, Siti Novia . *Prediksi Kecenderungan Politik dari Pengguna Twitter dengan Menggunakan Metode SVM*. Open library Telkom University 2016.
- [17] Tala, F. Z. (2003). A study of stemming effects on information retrieval in Bahasa Indonesia. Institute for Logic, Language and Computation, Universiteit van Amsterdam, The Netherlands.

Lampiran

1. Tweet Sentimen Ridwan Kamil



2. Tweet Setimen Deddy Mizwar



3. Tweet Sentimen Dedi Mulyadi

