

Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik di Facebook

Eko Budi Santoso¹, Aryo Nugroho²

¹Teknik Informatika, ²Fakultas Ilmu Komputer

^{1,2}Universitas Narotama Surabaya

Surabaya, Indonesia

e-mail: ¹ebud.santoso@gmail.com, ²aryo.nugroho@narotama.ac.id

Diajukan: 31 Mei 2019; Direvisi: 19 Agustus 2019; Diterima: 21 Agustus 2019

Abstrak

Analisis sentimen merupakan sebuah penelitian komputasional dari suatu pendapat sentimen dan luapan perasaan yang diekspresikan dalam bentuk tekstual. Facebook menjadi jejaring sosial yang terkenal di kalangan masyarakat Indonesia. Sebagai seorang tokoh politik yang mencalonkan sebagai presiden Indonesia, opini masyarakat sangat penting guna melihat dan mempertimbangkan popularitas dari seorang calon presiden. Tujuan penelitian ini adalah untuk mengklasifikasikan teks dengan baik terhadap sentimen positif dan negatif, sehingga informasi yang terdapat di dalamnya bisa diekstraksi serta dari data yang diamati dapat memberikan informasi yang berguna bagi pihak yang membutuhkannya. Beberapa tahapan dalam analisis sentimen antara lain pengumpulan data, preprocessing data, perhitungan skor sentimen, pelabelan sentimen dan pengklasifikasian data komentar memakai metode Naïve Bayes Classifier serta dilakukan asosiasi teks. Hasil dari penelitian berdasarkan data yang telah dikumpulkan sejak tanggal 17 April 2019 sampai 22 Mei 2019, Joko Widodo lebih unggul polaritas sentimen positif dari data sebanyak 5.000 komentar yang dipilih secara acak pada masing-masing calon presiden dan melalui tahap preprocessing yang menghasilkan polaritas sentimen. Joko Widodo diperoleh 85% sentimen positif, dan 15% untuk sentimen negatif. Sedangkan Prabowo Subianto diperoleh 76% sentimen positif, dan 24% untuk sentimen negatif. Metode klasifikasi Naïve Bayes Classifier terhadap penelitian ini memperoleh hasil akurasi 86,4%.

Kata kunci: Analisis Sentimen, Pemilihan Umum, Naïve Bayes Classifier, Asosiasi Text.

Abstract

Sentiment analysis is a computational study of sentiment opinions and feelings that are expressed in textual form. Facebook has become a popular social network among Indonesian people. As a political figure who is running for Indonesian president, public opinion is very important to see and consider the popularity of a presidential candidate. The purpose of this study is to classify texts well on positive and negative sentiments, so that the information contained in them can be extracted and from the observed data can provide useful information for those who need it. Some stages in sentiment analysis include data collection, data preprocessing, sentiment score calculation, sentiment labeling, and comment data classification using the Naïve Bayes Classifier method and text association. The results of the study are based on data collected from April 17 2019 to May 22 2019, Joko Widodo is superior to positive sentiment polarity from the data of 5,000 randomly selected comments on each presidential candidate and through the preprocessing stage which results in sentiment polarity. Joko Widodo obtained 85% positive sentiment, and 15% for negative sentiment. Whereas Prabowo Subianto obtained 76% positive sentiment, and 24% for negative sentiment. The Naïve Bayes Classifier classification method for this study obtained an accuracy of 86.4%.

Keywords: Sentiment Analysis, General Elections, Naïve Bayes Classifier, Text Association.

1. Pendahuluan

Indonesia merupakan negara yang mengikuti sistem demokrasi. Dalam hal ini didukung dengan adanya pemilu guna mencari seorang kepala negara. Di Indonesia, pemilu biasanya diselenggarakan secara serentak lima tahun sekali [1]. Pada tahun 2019 ini dilakukan pemilihan presiden Indonesia. Sebagai seorang politikus yang ingin maju mencalonkan sebagai presiden Indonesia, tentu akan mempertimbangkan

tingkat popularitas berdasarkan komentar masyarakat. Dalam mengekspresikan opini, masyarakat menggunakan jejaring sosial media yang terkenal saat ini yaitu *Facebook*. Pada awalnya *Facebook* hanya digunakan sebagai jejaring pertemanan saja, dengan seiring berjalannya waktu, *Facebook* telah mengalami pergeseran sebagai sarana interaksi untuk mempengaruhi orang lain [2] atau berpolitik dan berguna memberikan informasi kegiatan para tokoh politik atau publik. Setiap pengguna *Facebook* bebas mengunggah *posting*-an atau komentar tanpa ada batasan. Dengan keberadaan interaksi ini akan memberi efek interaksi sosial dan membentuk semacam jaringan pertemanan [3].

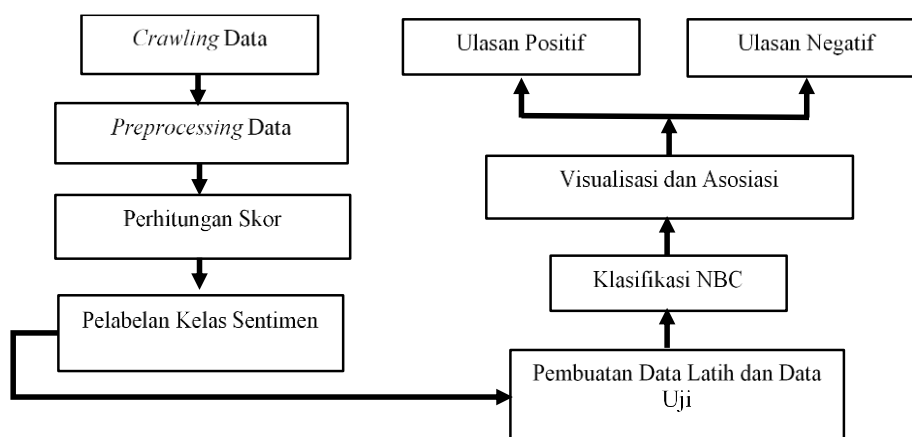
Pada umumnya *posting*-an di *Facebook* digunakan untuk mengunggah perihal tentang diri pengguna dan berbagi informasi, serta penyampaian sebuah berita. Isi dari *posting*-an juga dapat mengungkapkan sebuah perasaan dari pengguna. Oleh karena itu dibutuhkan sebuah analisis yang tepat, guna mengolah data teks sehingga menyuguhkan informasi yang berharga mengenai opini seseorang terhadap tokoh politik atau publik. Kumpulan komentar berupa teks tersebut dikumpulkan dan dapat diolah menggunakan analisis sentimen, yaitu mempelajari opini atau pendapat yang menyuarakan atau mengungkapkan persepsi yang mengandung sentimen positif atau negatif [4]. Pendapat yang dibutuhkan untuk melakukan analisis bersumber dari komentar halaman *fanpage* tokoh calon presiden Indonesia tahun 2019 di *Facebook*.

Pada penelitian yang dilakukan ini, data yang telah dikumpulkan kemudian akan masuk ke proses klasifikasi memakai metode klasifikasi yaitu *Naïve Bayes*. Metode ini merupakan sebuah metode klasifikasi yang memanfaatkan sebuah nilai dari probabilitas statistika sederhana yang mengasumsikan independen yang kuat dari aturan *Bayesian*. Metode analisis *Bayesian* ini melakukan sebuah analisis berdasarkan sebuah informasi *prior* dan informasi sampel. Gabungan dari informasi *prior* dengan informasi sampel tersebut diberi nama peluang *posterior*. Penerapan metode *Naïve Bayes Classifier* sendiri adalah menguji data *testing* dengan memanfaatkan data *training*.

Metode *Naïve Bayes Classifier* ini menghasilkan akurasi yang baik saat diterapkan dalam sebuah basis data yang banyak dan data yang berbeda-beda [5]. Hal ini dijelaskan oleh [6] di penelitian yang dilakukan, yaitu metode ini mempunyai beberapa nilai kemampuan lebih yaitu antara lain, cepat, sederhana, dan mempunyai akurasi tingkat tinggi. Harapan dari penelitian yang dilakukan ini bisa dan mampu mengklasifikasikan sebuah teks dengan bagus sehingga nanti sebuah informasi yang terkandung bisa diekstraksi serta informasi dari data yang disuguhkan bisa diamati dan dapat bermanfaat bagi pihak yang memerlukannya.

2. Metode Penelitian

Perencanaan alur analisis sentimen adalah sebuah gambaran umum yang menunjukkan alur dari penelitian yang sedang dilakukan. Alur kerja bisa dilihat pada Gambar 1 di bawah ini.



Gambar 1. Alur kerja penelitian.

Berdasarkan Gambar 1, alur yang dipakai pertama kali dalam penelitian ini adalah dengan cara *crawling* data komentar pada masing-masing akun *fanpage* calon presiden. Selanjutnya data yang diperoleh masuk pada tahap *preprocessing* antara lain yaitu tahap *cleansing*, *case folding*, *tokenizing*, dan *filtering*. Berikutnya dari hasil *preprocessing* dilakukan perhitungan skor dan pelabelan kelas sentimen berdasarkan kamus *lexicon*. Kemudian data dibagi menjadi data latih dan data uji, data tersebut diklasifikasikan

menggunakan metode *Naïve Bayes Classifier* serta dilakukan visualisasi serta asosiasi teks ulasan kata positif dan kata negatif.

2.1. Pengumpulan Data

Pengambilan data yang dilakukan akan menggunakan metode *crawling* memakai *tool Facepager* dengan memanfaatkan *access token* dari *Facebook*. Metode ini memungkinkan untuk memanggil data berupa status atau komentar *Facebook* yang diinginkan pada halaman *fanpage Facebook* terhadap calon presiden Indonesia tahun 2019 sebagai sumber data. Pada tahap pengumpulan data perlu tingkat ketelitian dan kesabaran. Karena data yang terkumpul harus memenuhi setiap kriteria yang dibutuhkan seperti menghindari komentar berisi *emoticon* dan karakter lainnya. Untuk perihal jumlah data yang diperlukan harus memenuhi target yang telah ditentukan. Pengumpulan data sendiri dilakukan pada tanggal 17 April 2019 sampai 22 Mei 2019. Data yang dikumpulkan adalah data komentar dari halaman *fanpage* Joko Widodo dan Prabowo Subianto, selanjutnya dilakukan proses pengolahan data sampai menghasilkan *dataset* yang siap diklasifikasi.

2.2. Data Correction

Konsep utama dari sebuah data adalah supaya bisa dimanfaatkan secara penuh sebagai bukti yang *valid* pada suatu peristiwa. Oleh sebab itu tingkat akurasi dan integritas dalam proses pengelolaan harus diperhatikan. Jika dua komponen tersebut sudah terpenuhi maka data dapat dipakai guna meningkatkan kualitas dari proses pengambilan sebuah keputusan. Dalam melakukan pengecekan data mula-mula definisikan terlebih dahulu aturan yang harus dipenuhi sebuah data. Misalnya rentang nilai, jumlah kolom, dan kategori data. Semua ini dilakukan untuk memudahkan proses pembuangan data yang tidak dibutuhkan.

2.3. Preprocessing

Data komentar calon presiden yang diperoleh dari proses *crawling* pada situs jejaring sosial *Facebook* dan belum bisa dipakai karena disebabkan masih mempunyai bentuk teks atau kalimat yang tidak beraturan atau tidak terstruktur di mana memiliki banyak *noise*. Oleh sebab itu, pada proses ini diperlukan perubahan bentuk dari data yang tidak terstruktur menjadi data yang terstruktur yang mempunyai nilai numerik [7]. Setelah data menjadi terstruktur dan mempunyai nilai numerik maka data bisa diproses untuk mengekstrak informasi dari dokumen-dokumen teks yang bisa digunakan untuk kepentingan analisis berbagai bidang multidisipliner yaitu klasifikasi, klusterisasi, informasi temu kembali, visualisasi atau berbagai analisis teks yang lainnya [8]. Adapun tahapan *preprocessing* data adalah sebagai berikut:

- *Cleansing*, yaitu proses di mana menghapus atribut atau tanda baca yang tidak diperlukan pada proses analisis.
- *Case folding*, yakni mengubah semua huruf besar menjadi *lowercase* atau huruf kecil.
- *Tokenizing* merupakan proses pemisahan kata berdasarkan tiap kata yang menyusunnya menjadi potongan tunggal.
- *filtering*, yaitu tahap menghapus kumpulan dari kata yang tidak berpengaruh tanpa mengurangi informasi dari dokumen tersebut dan keberadaannya sering muncul. Dengan memanfaatkan *stopword* berbahasa Indonesia yang telah disusun dan dilakukan penambahan kata secara manual karena disebabkan ada dari beberapa kata yang belum terdapat dalam *stopword*.

2.4. Perhitungan Skor Sentimen

Berdasarkan teks komentar “Jalan tol yang sudah dikerjakan sangat bagus dan indah, tapi untuk pembayaran tolnya sangat mahal, terdapat 1 kata negatif dan 2 kata positif yang terdeteksi pada kamus *lexicon*, yaitu “bagus” dan “indah” sebagai kata positif, untuk kata negatif yaitu “mahal” sebagai kata negatif. Adapun rumus yang digunakan dalam proses perhitungan skor sentimen adalah sebagai berikut:

$$Skor = (\sum \text{kata positif}) - (\sum \text{kata negatif}) \tag{1}$$

Tabel 1. Perhitungan Skor.

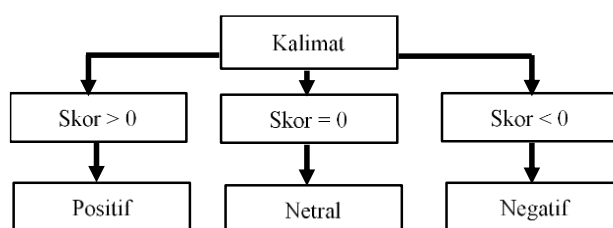
Komentar	Jumlah Kata Positif	Jumlah Kata Negatif
Jalan tol yang sudah dikerjakan sangat bagus dan indah tapi untuk pembayaran tolnya sangat mahal	bagus indah	mahal
Total Jumlah Kata	2	1

Maka, diperoleh hasil dari perhitungan:
 Skor = (Jumlah kata positif) – (Jumlah kata negatif)
 Skor = 2 – 1 = 1

Nilai akhir yang diperoleh dari perhitungan menghasilkan skor 1 atau > 0, maka diidentifikasi kata positif.

2.5. Pelabelan Kelas Sentimen

Pada bagian ini juga merupakan salah satu proses untuk mendapatkan hasil representasi *corpus* yang diharapkan. Proses pelabelan sendiri dilakukan dengan cara menghitung sebuah nilai secara otomatis dan pelabelan sentimen memakai kamus *lexicon* serta dengan cara manual. Jika dalam kalimat mempunyai skor > 0 akan diklasifikasikan dalam kelas positif, jika kalimat mempunyai skor = 0 akan diklasifikasikan dalam kelas netral, sedangkan jika kalimat mempunyai skor < 0 diklasifikasikan dalam kelas negatif [9]. Adapun proses pelabelan dapat dilihat pada Gambar 2 di bawah ini:



Gambar 2. Pelabelan kelas sentimen.

2.6. Klasifikasi Naïve Bayes Classifier

Metode klasifikasi ini memakai konsep sebuah peluang dalam menentukan kelas dalam suatu dokumen yang mengasumsikan bahwa kemunculan kata tidak mempengaruhi kemunculan kata yang lain dalam sebuah dokumen dan ketidakmunculan kata tidak mempengaruhi ketidakmunculan kata yang lain dalam sebuah dokumen [10]. *Naïve Bayes* adalah suatu teknik prediksi yang berdasar terhadap penerapan *teorema Bayes* dengan pengertian independensi yang kuat [11]. *Naïve Bayes Classifier* memberi pengertian ada atau tidak adanya suatu fitur tertentu pada sebuah kelas pada suatu dokumen tidak mempengaruhi keberadaan fitur yang lain.

2.7. Asosiasi Teks

Dalam penelitian oleh [12] disebutkan bahwa istilah asosiasi sering diasumsikan tentang hubungan antara dua variabel atau lebih yang memiliki sifat kualitatif. Analisis korelasi sendiri memiliki tujuan guna mencari hubungan variabel bebas (*X*) dengan variabel terikat (*Y*), yang mempunyai ketentuan data dan syarat tertentu [13].

Persamaan (2) digunakan untuk menghitung nilai korelasi dan persamaan (3) untuk mendapatkan *R Square*. Penelitian ini memakai sebuah pendekatan asosiasi guna memperoleh hubungan antar komentar dari pengguna sehingga mendapat informasi yang bisa dibuat sebagai bahan rujukan dalam proses evaluasi untuk memperbaiki elektabilitas calon presiden Indonesia tahun 2019.

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{n \cdot \sum X^2 - (\sum X)^2} \cdot \sqrt{n \cdot \sum Y^2 - (\sum Y)^2}} \tag{2}$$

$$R = r^2 \tag{3}$$

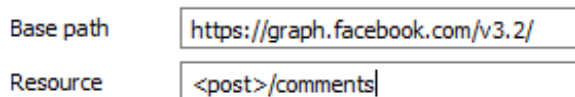
3. Hasil dan Pembahasan

3.1. Data Crawling

Pengumpulan data dalam penelitian ini dilakukan dengan metode *crawling* menggunakan *tool Facepager*. Di mana proses pengambilan data membutuhkan *aces token* dari *Facebook*. Cara menjalankan *tool* ini adalah dengan memasukan *aces token* dan *login* ke akun *Facebook* yang sudah dibuat dan memasukkan nama halaman *fanpage* untuk mengambil data status dan komentar dari halaman tersebut. Berikut perintah mengambil data status dan komentar pada halaman *Facebook* calon presiden Indonesia tahun 2019 pada Gambar 3 dan 4 di bawah ini.



Gambar 3. Pengambilan data status pada halaman Facebook.



Gambar 4. Pengambilan data komentar pada halaman Facebook.

Pengumpulan data status dalam penelitian ini diambil sejak tanggal 17 April 2019 sampai 22 Mei 2019 dari akun Facebook Joko Widodo dan Prabowo Subianto. Berikut adalah jumlah data status dari masing-masing halaman Facebook calon presiden Indonesia tahun 2019 yang menyatakan pada halaman Facebook Joko Widodo lebih banyak *posting*-an yang berjumlah 40 status, sedangkan Prabowo Subianto hanya 12 status. Untuk data komentar yang dipilih sebanyak 5.000 data komentar dari para pengguna jejaring sosial Facebook terhadap masing-masing calon presiden secara *random*. Berikut hasil pengumpulan data dapat dilihat pada Tabel 2 di bawah ini.

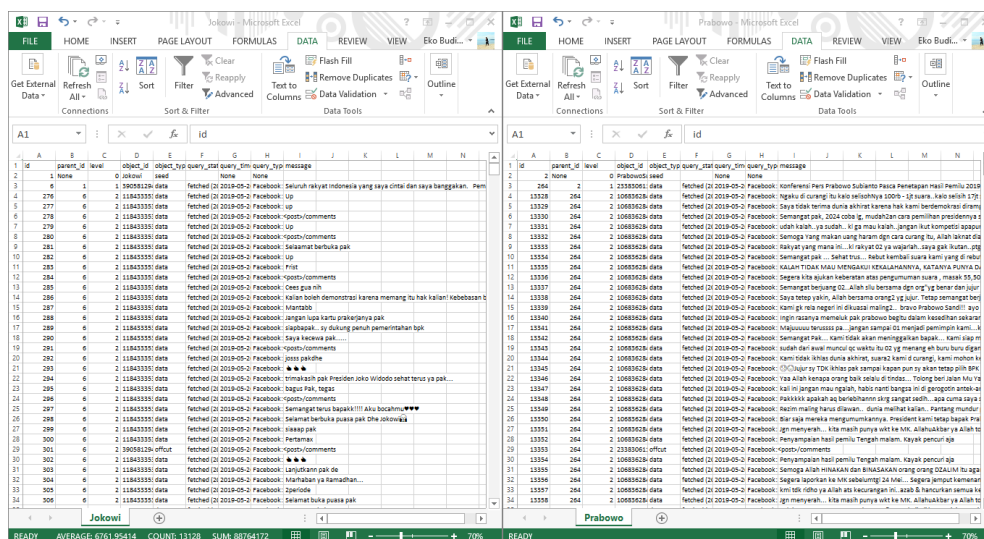
Tabel 2. Hasil pengumpulan data.

Nama	Jumlah Status	Jumlah Komentar
Joko Widodo	40	5.000
Prabowo Subianto	12	5.000

3.2. Data Correction

Format data yang diperoleh dari *tool* Facepacer adalah berformat CSV (*Comma Separated Values*), yaitu format data di mana setiap *record* dipisah oleh tanda koma (,) dan tanda petik dua (") atau pembatas (*delimiter*) lainnya yang umum dipakai. Dalam aplikasi ini pembatas yang dipakai adalah koma (,). Oleh sebab itu sebelum melanjutkan ke tahap berikutnya, data yang masih mentah ini harus dilakukan pengecekan kelengkapannya.

Langkah pertama yang dilakukan pada tahap *correction* ini adalah dengan *import file CSV* ke dalam *MS.Excel*, dengan ini format data akan berubah menjadi *xlsx* dari format *csv*, sehingga proses pengoreksian bisa dilakukan dengan mudah. Setelah data sudah diimpor, maka data akan melalui tahap selanjutnya yaitu pemuangan atribut yang tidak dibutuhkan serta penghapusan data yang terdeteksi duplikasi. Pada penelitian ini penulis hanya memakai data komentar yang berisi data teks berbahasa Indonesia. Pada Gambar 5 di bawah ini adalah hasil dari data *correction*.



Gambar 5. Hasil data correction.

3.3. Preprocessing Data

Data yang diproses dalam tahap *preprocessing* hanya data komentar yang telah berhasil dikumpulkan dan tela melalui tahap data *correction*. Adapun langkah-langkah *preprocessing* yang digunakan adalah *tahap cleansing, case folding, tokenizing, dan filtering*. Pada Tabel 3 di bawah ini menunjukkan proses dari tahap *preprocessing* data.

Tabel 3. Proses *preprocessing* data.

Tahap	Input	Output
<i>Cleansing</i>	Hoax.. Mana bisa Jam 10 pagi di Manado sekarang sudah di Makassar.. Calon.Presiden kok tukang bohong tukang tipu?????	Hoax Mana bisa Jam pagi di Manado sekarang sudah di Makassar Calon Presiden koq tukang bohong tukang tipu
<i>Case folding</i>	Hoax Mana bisa Jam pagi di Manado sekarang sudah di Makassar Calon Presiden kok tukang bohong tukang tipu	hoax mana bisa jam pagi di manado sekarang sudah di makassar calon presiden kok tukang bohong tukang tipu
<i>Tokenizing</i>	hoax mana bisa jam pagi di manado sekarang sudah di makassar calon presiden kok tukang bohong tukang tipu	“hoax” “mana” “bisa” “jam” “pagi” “di” “manado” “sekarang” “sudah” “di” “makassar” “calon” “presiden” “kok” “tukang” “bohong” “tukang” “tipu”
<i>Filtering</i>	“hoax” “mana” “bisa” “jam” “pagi” “di” “manado” “sekarang” “sudah” “di” “makassar” “calon” “presiden” “kok” “tukang” “bohong” “tukang” “tipu”	hoax jam pagi manado sudah makassar calon kok tukang bohong tukang tipu

3.4. Pelabelan Kelas Sentimen

Proses pelabelan pada dasarnya terbagi menjadi tiga kategori kelas sentimen, yakni sentimen negatif, sentimen positif dan sentimen netral dengan melakukan perhitungan nilai. Suatu dokumen dapat masuk ke dalam kelas positif atau negatif dengan memanfaatkan kumpulan kata berbahasa Indonesia atau sebuah kamus yang terdiri dari kumpulan kata positif dan kumpulan kata negatif. Selanjutnya dilakukan proses pelabelan menggunakan aplikasi *R* secara otomatis dengan cara menghitung skor jumlah kata positif yang ada pada kalimat akan dikurangi dengan skor jumlah kata negatif dalam suatu kalimat komentar. Gambar 6 di bawah merupakan *Script R programming* untuk pelabelan kelas sentimen.

```
list.kata= str_split(kalimat, '\\s+')
kata2= unlist(list.kata)
positif.matches= match(kata2, kata.positif)
negatif.matches= match(kata2, kata.negatif)
positif.matches= !is.na(positif.matches)
negatif.matches= !is.na(negatif.matches)
score= sum(positif.matches) - (sum(negatif.matches))
return(score)
}, kata.positif, kata.negatif, .progress=.progress )
scores.df= data.frame(score=scores, text=docs)
return(scores.df)
}

hasil= score.sentiment(docs$text, kata.positif, kata.negatif)
View(hasil)

hasil$klasifikasi<-ifelse(hasil$score<0,"Negatif","Positif")
hasil$klasifikasi
View(hasil)
```

Gambar 6. *Script* pelabelan kelas sentimen menggunakan aplikasi *R*.

Perbandingan hasil jumlah data dari proses pelabelan kelas sentimen menggunakan bahasa *R programming*, sehingga diperoleh data komentar yang sudah terlabeli secara otomatis pada masing-masing calon presiden yang ditunjukkan pada Tabel 4 dan 5 sebagai berikut.

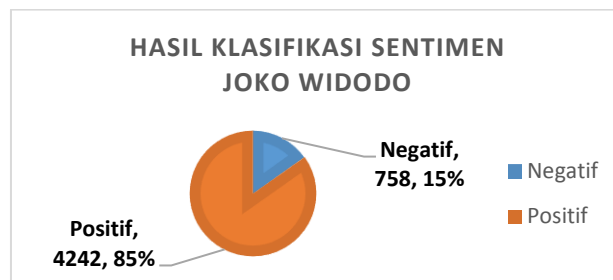
Tabel 4. Perbandingan jumlah data pada kelas sentimen terhadap Joko Widodo.

Nama	Sentimen	Jumlah Komentar
Joko Widodo	Positif	2157
	Negatif	758
	Netral	2085

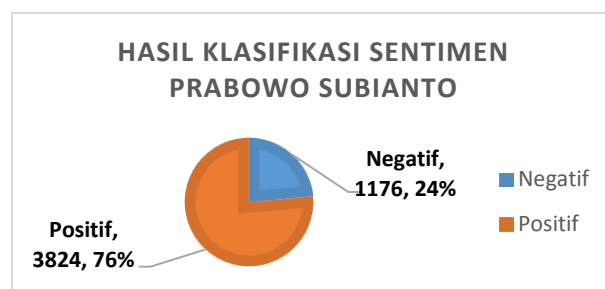
Tabel 5. Perbandingan jumlah data pada kelas sentimen terhadap Prabowo Subianto.

Nama	Sentimen	Jumlah Komentar
Prabowo Subianto	Positif	3069
	Negatif	1176
	Netral	755

Selanjutnya dilakukan proses reduksi pada kelas dengan memasukkan kategori kelas sentimen netral ke sentimen positif yang dilakukan secara manual. Berikut tampilan presentasi kelas sentimen dari masing-masing calon presiden Indonesia tahun 2019.



Gambar 7. Hasil kelas sentimen terhadap Joko Widodo.



Gambar 8. Hasil kelas sentimen terhadap Prabowo Subianto.

Berdasarkan Gambar 7 dan 8, terlihat jumlah data komentar terhadap calon presiden Indonesia tahun 2019 masing-masing 5.000 data komentar pada Joko Widodo dan 5.000 data komentar pada Prabowo Subianto. Dengan presentasi Joko Widodo mendapatkan 85% sentimen positif dan 15% data negatif, sedangkan Prabowo Subianto mendapatkan 76% sentimen positif dan 24% data negatif. Dalam hal ini maka calon presiden Joko Widodo lebih unggul dalam polaritas sentimen dibanding dengan Prabowo Subianto.

3.5. Klasifikasi Dengan Naïve Bayes Classifier

Pada tahap proses klasifikasi dalam penelitian yang dilakukan ini adalah dengan membangun sebuah *machine learning* dengan memakai data latih serta data uji dari semua data terhadap calon presiden secara *random* dengan dilakukan iterasi terhadap *dataset* sebagai *cross validation* guna memperoleh nilai akurasi prediksi. Berikut adalah tampilan hasil tahap klasifikasi *Naïve Bayes Classifier* dengan menggunakan *R Programming*.

```

Accuracy : 0.864
95% CI : (0.8482, 0.8787)
No Information Rate : 0.82
P-Value [Acc > NIR] : 6.889e-08

Kappa : 0.5173

Mcnemar's Test P-Value : 0.01292
    
```

Gambar 9. Hasil klasifikasi menggunakan *Naïve Bayes Classifier*.

Gambar 9 menunjukkan hasil akurasi yang diperoleh menggunakan *Naïve Bayes Classifier* (NBC) dengan perbandingan data latih sebesar 80% dan data uji sebesar 20% dari 10.000 data komentar dengan akurasi sebesar 86,4%. Dengan ini dapat dikatakan bahwa metode *Naïve Bayes Classifier* ini baik dipakai dalam proses pengklasifikasian komentar dengan teks bahasa Indonesia terhadap calon presiden Indonesia tahun 2019. Penulis juga menggunakan metode *confusion matrix* dalam proses evaluasi mencari nilai akurasi, nilai presisi dan nilai *recall*. Metode ini adalah salah satu *tools* penting dalam metode visualisasi yang dipakai pada *machine learning* yang biasanya memuat dua kategori atau lebih [14]. Pada matriks

bagian baris merupakan contoh jumlah data uji untuk kelas sebenarnya sedangkan untuk bagian kolom merupakan kelas yang diprediksi.

Tabel 6. Hasil *confusion matrix*.

	Positif	Negatif
Positif	203	115
Negatif	157	1525

Pada Tabel 6 menunjukkan hasil dari prediksi 318 data komentar yang masuk kategori sentimen positif dan negatif, didapatkan 203 data komentar yang terklasifikasi dengan benar dan 115 data komentar yang tidak terklasifikasi dengan benar. Sedangkan dari 1.682 data komentar yang masuk kategori sentimen positif dan negatif, didapatkan 1.525 data komentar terklasifikasi dengan benar dan 157 data komentar yang tidak terklasifikasi dengan benar. Berikut perhitungan presisi dan *recall* pada persamaan (4) dan (5) di bawah ini.

$$precision = \frac{TP}{TP + FP} * 100\% = \frac{203}{203 + 115} * 100\% = 63,8\% \quad (4)$$

$$recall = \frac{TP}{TP + FN} * 100\% = \frac{203}{203 + 157} * 100\% = 56,3\% \quad (5)$$

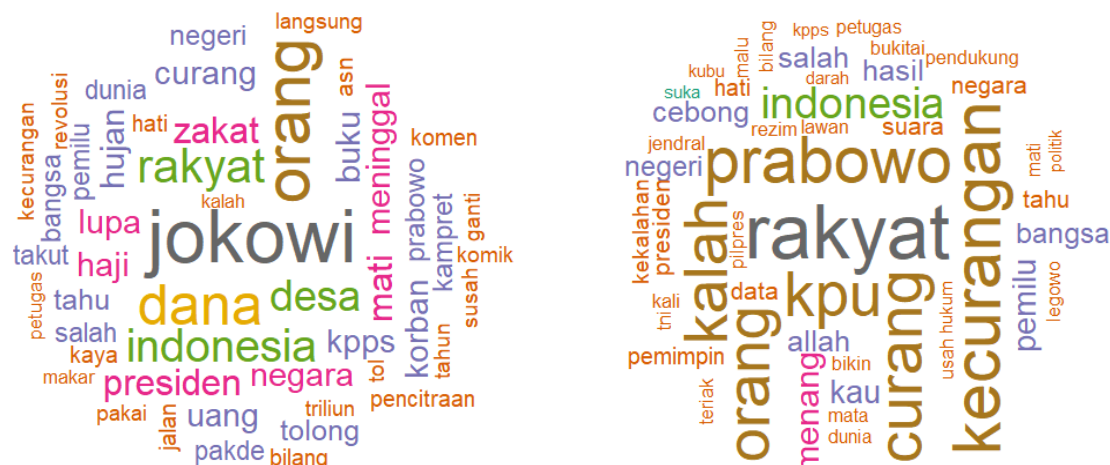
Nilai presisi diperoleh sebesar 63,8% dan *recall* 56,3%. Kemudian untuk *confusion matrix* menghasilkan tingkat akurasi rata-rata sebesar 74.6%. yang artinya dari data 2.000 data komentar yang diuji, terdapat 1.728 data komentar yang benar diklasifikasi oleh model *Naive Bayes Classifier*.

3.6. Visualisasi dan Asosiasi

Wordcloud merupakan sistem yang mampu memvisualisasikan kata-kata dengan menyoroti frekuensi kata yang digunakan dalam suatu teks tertulis dan sebagai alat bantu untuk melakukan analisis guna memperoleh wawasan holistik terhadap para calon presiden Indonesia tahun 2019.



Gambar 10. Kata positif yang paling banyak muncul terhadap Joko Widodo dan Prabowo Subianto.



Gambar 11. Kata negatif yang paling banyak muncul terhadap Joko Widodo dan Prabowo Subianto.

Dapat dilihat pada Gambar 10 dan 11, hasil visualisasi *wordcloud* bisa dilihat kata-kata positif dan negatif yang sering dipakai dalam komentar. Besar ukuran kata pada *wordcloud* mengartikan frekuensi kemunculan kata tersebut sering dipakai dalam memberikan komentar, yang berarti para pengguna Facebook sering memakai kata tersebut sebagai topik pembicaraan dengan penilaian positif dan negatif dalam komentar. Selanjutnya, pencarian asosiasi antar kata yang berhubungan dengan kata masyarakat diperoleh hasil sebagai berikut:

Tabel 7. Asosiasi kata terhadap Joko Widodo dan Prabowo Subianto.

Asosiasi Kata Positif Joko Widodo	Asosiasi Kata Negatif Joko Widodo	Asosiasi Kata Positif Prabowo Subianto	Asosiasi Kata Negatif Prabowo Subianto		
Upaya	0,62 Pemerintahan	0,69	Sistem	0,62 Bodohi	0,66
Mental	0,60 Pembangunan	0,58	Berkomitmen	0,60 Kelayakan	0,65
Kondisi	0,56 Kelompok	0,56	Evaluasi	0,56 Diusung	0,65
Berlanjut	0,53 Tanah	0,50	Kritis	0,53 Komisioner	0,65
Berkonsultasi	0,53 Berhutang	0,50	Kemanusiaan	0,53 Fraksi	0,65

Berdasarkan Tabel 7 menunjukkan kata-kata yang berasosiasi dengan kata “masyarakat” pada komentar positif memberikan informasi topik perbincangan terhadap Joko Widodo antara lain tentang upaya, mental, kondisi, berlanjut, dan berkonsultasi. Sedangkan terhadap Prabowo Subianto antara lain tentang sistem, berkomitmen, evaluasi, kritis, dan kemanusiaan berdasarkan 5 kata teratas, dan untuk komentar negatif sendiri memberikan informasi keluhan terhadap Joko Widodo antara lain tentang pemerintahan, pembangunan, kelompok, tanah, dan berhutang. Sedangkan terhadap Prabowo Subianto antara lain tentang bodohi, kelayakan, diusung, komisioner, dan fraksi berdasarkan 5 kata teratas.

4. Kesimpulan

Dari analisis yang telah dilakukan, didapatkan beberapa kesimpulan antara lain jumlah data *posting*-an Joko Widodo berjumlah 40 status, sedangkan Prabowo Subianto berjumlah 12 status sejak tanggal 17 April 2019 sampai 22 Mei 2019. Untuk pengamatan polaritas sentimen publik sendiri menunjukkan calon presiden Joko Widodo mendapatkan 4.242 data komentar untuk kelas sentimen positif, dan 758 data komentar kelas sentimen negatif. Sedangkan Prabowo Subianto memperoleh 3.824 data komentar untuk kelas sentimen positif, dan 1.176 data komentar kelas sentimen negatif dari masing-masing calon presiden dengan 5.000 data komentar. Dengan ini Joko Widodo memiliki polaritas sentimen yang lebih unggul dari Prabowo Subianto.

Informasi yang diperoleh dari hasil asosiasi teks secara umum menunjukkan bahwa pada Joko Widodo, kata-kata yang berasosiasi dengan kata “masyarakat” banyak membicarakan tentang kata upaya, mental, kondisi, berlanjut, dan berkonsultasi untuk 5 kata teratas sentimen positif, dan kata pemerintahan, pembangunan, kelompok, tanah, dan berhutang untuk 5 kata teratas sentimen negatif. Sedangkan pada Prabowo Subianto, kata-kata yang berasosiasi dengan kata “masyarakat” banyak membicarakan tentang

kata sistem, berkomitmen, evaluasi, kritis, dan kemanusiaan untuk 5 kata teratas sentimen positif, dan kata bodohi, kelayakan, diusung, komisioner, dan fraksi untuk 5 kata teratas sentimen negatif.

Daftar Pustaka

- [1] R. Hidayatillah, M. Hakam, and A. Nugroho, "Levels of Political Participation Based on Naive Bayes Classifier," vol. 13, no. 1, pp. 73–82, 2019.
- [2] A. Nugroho, M. Noor, A. Azam, and S. Anam, "Pembangunan Aplikasi Community Messenger Sebagai Alat Interaksi Di Kalangan Generasi C," vol. 2, no. 2, 2016.
- [3] A. Nugroho, S. Supeno, and M. H. Purnomo, "Visualizing Interaction in Catfiz Indonesian Messenger Using Graph Coloring Visualizing Interaction in Catfiz Indonesian Messenger Using Graph Coloring," 2015.
- [4] B. Liu, *Sentiment Analysis(Introduction and Survey) and Opinion Mining*. 2012.
- [5] D. T. Larose, "Naive bayes estimation and bayesian networks," in *Data Mining Methods and Models*, 2006.
- [6] R. Mccue, "A Comparison of the Accuracy of Support Vector Machine and Naive Bayes Algorithms In Spam Classification," *Univ. Calif. St. Cruz*, pp. 1–17, 2009.
- [7] S. M. Weiss, N. Indurkha, and T. Zhang, *Fundamentals of Predictive Text Mining (Texts in Computer Science)*. 2010.
- [8] A. Akilan, "Text mining: Challenges and future directions," *2nd Int. Conf. Electron. Commun. Syst. ICECS 2015*, pp. 1679–1683, 2015.
- [9] E. W. Pamungkas and D. G. P. Putri, "An experimental study of lexicon-based sentiment analysis on Bahasa Indonesia," *Proc. - 2016 6th Int. Annu. Eng. Semin. Ina. 2016*, pp. 28–31, 2017.
- [10] B. Rianto, "Implementasi dan perbandingan metode prapemrosesan pada analisis sentimen gubernur DKI Jakarta menggunakan metode support vector machine dan naive bayes." Skripsi : Program Studi Ilmu Komputer FMIPA UGM Yogyakarta, 2016.
- [11] N. W. S. Saraswati, "Naïve Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis," *Semin. Nas. Sist. Inf. Indones.*, pp. 586–591, 2013.
- [12] M. N. Ulwan, "Pattern Recognition Pada Unstructured Data Teks Menggunakan Support Vector Machine Dan Association." Skripsi: Program Studi Statistika Universitas Islam Indonesia, 2016.
- [13] Bustami, D. Abdullah, and Fadlisyah, *Statistika : Terapannya di Informatika*. 2014.
- [14] C. D. Manning, P. Raghavan, and H. Schütze, "An Introduction to Information Retrieval," *Cambridge Univ. Press*, vol. 38, no. c, pp. 156–164, 2009.