

# **Analisis Sentimen Pemilihan Presiden Indonesia Tahun 2019 Di Twitter Berdasarkan *Geolocation* Menggunakan Metode *Naïve Bayesian Classification***

**Wiranto Horsen Silitonga<sup>1</sup>, Jay Idoan Sihotang\*<sup>2</sup>**

<sup>1</sup>PT. Bank OCBC NISP, Indonesia

<sup>2</sup>Fakultas Teknologi Informasi, Universitas Advent Indonesia  
e-mail: <sup>1</sup>wirantohorsen@gmail.com, \*<sup>2</sup>jay.sihotang@unai.edu

## **Abstrak**

Pemilihan Presiden Indonesia 2019 ramai diperbincangkan di dunia nyata maupun dunia maya, khususnya di media sosial Twitter. Semua orang bebas berpendapat tentang pasangan calon Presiden Indonesia 2019 tersebut. Sehingga memunculkan banyak opini, tidak hanya opini yang positif atau netral, ada pula opini negatif. Media sosial khususnya Twitter sekarang ini menjadi salah satu tempat promosi atau kampanye yang efektif dan efisien untuk menggait para pendukung. Dalam hal ini peneliti akan melakukan riset terhadap tokoh publik yang mencalonkan diri menjadi Presiden Indonesia. Metode penelitian yang digunakan dalam riset kali ini adalah algoritma klasifikasi *Naïve Bayesian Classifier*. Data yang digunakan adalah *tweet* berbahasa Indonesia dengan kata kunci Jokowi (*#Jokowi2Periode*) dan Prabowo (*#PrabowoSandi*) sebanyak 1009 data *tweet* selama 5 bulan dimulai dari 1 September 2019 sampai 31 Januari 2019. Yang di mana data *tweet* tersebut diambil dari empat daerah terbesar di Indonesia, yaitu Jakarta, Bandung, Medan, dan Surabaya. Setiap data akan diambil secara manual menggunakan *Geolocation* API yang telah di sediakan oleh Twitter melalui *Twitter search*. Hasil dari klasifikasi menggunakan algoritma *Naïve Bayesian Classifier* didapat 839 *tweet* positif, 32 *tweet* negatif, dan 67 *tweet* netral dari 938 *tweet* keseluruhan, atau dalam bentuk persentase ada 90% merupakan sentimen positif, 3% sentimen negatif, dan 7% sentimen netral terhadap bapak Joko Widodo. Dan 56 *tweet* positif, 6 *tweet* negatif, dan 8 *tweet* netral dari 70 *tweet* keseluruhan, atau dalam bentuk persentase ada 80% merupakan sentimen positif, 9% sentimen negatif, dan 11% sentimen netral terhadap bapak Prabowo. Tingkat akurasi yang dihasilkan dari algoritma *Naïve Bayesian Classifier* sendiri terhadap penelitian ini sebesar 77,62%.

**Kata Kunci:** Analisa Sentimen, Pemilihan Presiden Indonesia, Jokowi, Prabowo, *Geolocation*, *Datamining*, *Naïve Bayesian Classifier*, *Multinomial Naïve Bayes*

## ***Sentiment Analysis of Indonesia Presidency Election 2019 on Twitter Based on Geolocation Using Naïve Bayesian Classification Method***

### **Abstract**

*2019 Indonesian Presidential Election is crowded to be discussed in the real world and also cyberspace, specifically on Twitter. Everyone is free to agree on the 2019 Indonesian Presidential candidate pair. Opinion raises many opinions, not only positive or neutral opinions but there are also negative opinions. Twitter's is now one of the most effective and efficient promotional or campaign venues to attract supporters. In this case, the researcher will conduct research on community leaders who are running for the presidency of Indonesia. The research method used in this study is the Naïve Bayesian Classifier classification algorithm. The data used are Indonesian tweets with Jokowi (*#Jokowi2Periode*) and Prabowo (*#PrabowoSandi*) keywords totaling 1009 data tweets for 5 months starting from September 1, 2019 to 31 January 1, 2019. Indonesia, namely Jakarta, Bandung, Medan, and Surabaya. Each data will be taken manually by using the Geolocation API that has been provided by Twitter via a Twitter search. The results*

of the classification using the Naïve Bayesian Classifier algorithm received 839 positive tweets, 32 negative tweets, and 67 neutral tweets from 938 overall tweets, or in the form of a percentage, there were 90% containing positive sentiment, 3% negative, and 7% negative sentiment towards Mr. Joko Widodo. And 56 positive tweets, 6 negative tweets, and 8 neutral tweets from 70 overall tweets, or in the form of the percentage there are 80% positive sentiments, 9% negative sentiments, and 11% neutral sentiments towards Mr. Prabowo. The level of accuracy generated from the Naïve Bayesian Classifier algorithm itself for this study amounted to 77.62%.

**Keywords:** Sentiment Analysis, Indonesia Presidency Election, Jokowi, Prabowo, Geolocation, Datamining, Naïve Bayesian Classifier, Multinomial Naïve Bayes

## 1. Pendahuluan

Sistem demokrasi adalah suatu sistem yang berlaku pada negara Indonesia. Hal tersebut dibuktikan dengan diadakannya pemilihan umum pada setiap ajang penentuan pemimpin yang baru. Baik itu presiden, wakil presiden, ataupun lembaga perwakilan rakyat. Pada setiap ajang pemilihan umum tersebut, setiap calon tokoh politik akan melihat atau mempertimbangkan popularitas mereka dari setiap opini masyarakat. Dikarenakan perkembangan zaman yang sudah begitu pesat hingga merambah sampai perkembangan teknologi komunikasi saat ini, mengubah kecenderungan masyarakat dalam mengekspresikan opininya pada sebuah jejaring sosial. Dan Twitter adalah salah satu jejaring sosial yang sangat populer di kalangan pengguna internet saat ini [1][2].

Dan isi pada Twitter sendiri yang disebut *tweet* berisikan tentang apa yang sedang dilakukan atau dirasakan, percakapan, berbagi informasi, dan pelaporan suatu berita. Pada umumnya sebuah *tweet* hanya untuk melakukan *posting* hal mengenai diri pengguna dan berbagi informasi. Isi *tweet* juga dapat mengekspresikan perasaan atau *mood* pengguna, hal ini bersifat penilaian subjektif atau opini. Opini melalui *tweet* inilah yang dapat dimanfaatkan peneliti untuk melakukan analisis sentimen masyarakat terhadap tokoh politik yang akan maju sebagai calon presiden Indonesia tahun 2019.

Dengan banyaknya *tweet* yang beredar di Twitter yang membahas mengenai Pemilihan Presiden, peneliti termotivasi untuk melihat Analisa sentimen dari *tweet* yang ada di Twitter. Analisis sentimen dan *opinion mining* adalah bidang studi yang menganalisis pendapat seseorang, sentimen seseorang, evaluasi seseorang, sikap seseorang dan emosi seseorang ke dalam bahasa tertulis. Teknik analisis sentimen dapat mendukung banyak keputusan dalam banyak skenario. Penelitian ini menggunakan tiga *class attribute*, yaitu positif, netral dan negatif, karena di internet komentar yang muncul dapat berupa komentar positif, netral dan negatif [3].

Data terhadap penelitian ini diperoleh melalui *tweet search* pada aplikasi Twitter dengan *hashtag* dan *geolocation* yang telah ditentukan. Sehingga diperolehlah sebuah dokumen baru yang digolongkan menjadi 3 bagian polaritas yaitu, positif, negatif, dan netral. Di mana hal ini dilakukan untuk diperolehnya data yang akurat. Kumpulan data yang sudah diperoleh akan melewati tahap *PreProcessing*. Hasil dari tahap *PreProcessing* tersebut akan diklasifikasikan dengan metode *Naïve Bayes Classifier* [4].

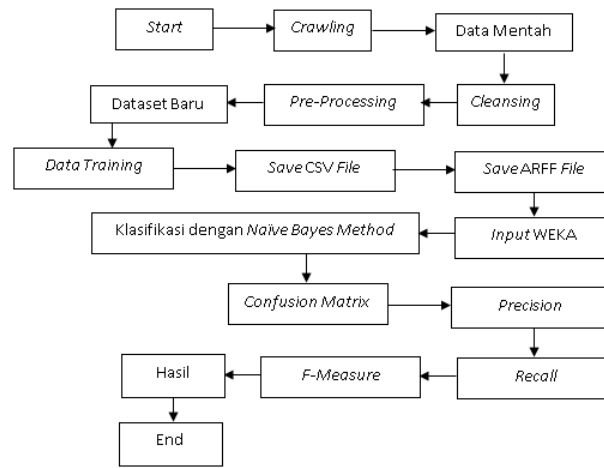
## 2. Metode Penelitian

### Pengumpulan Data

*Dataset* yang akan digunakan dalam penelitian *tweet* bahasa Indonesia yang merupakan hasil pencarian berdasarkan *geolocation* empat daerah yaitu Jakarta, Bandung, Surabaya, dan Medan terhadap tagar tokoh politik yang akan maju sebagai calon presiden Indonesia 2019. Beberapa tagar yang diambil di antaranya, *#jokowi2periode* (Joko Widodo) dan *#prabowosandi* (Prabowo Subianto). *Dataset* didapatkan dengan cara manual memanfaatkan API dari Twitter *Search* kemudian disimpan dalam format Ms. Excel. Pengambilan data *tweet* dilakukan pada 1 September 2018 hingga 31 Januari 2019.

## Implementasi

Implementasi penelitian ini dilakukan dengan langkah kerja seperti Gambar.1.



**Gambar 1** Diagram Blok

Proses awal pada tahapan implementasi adalah pengumpulan *dataset* yang akan digunakan baik untuk *testing* maupun *training*. Pengumpulan *dataset* dengan cara memanfaatkan API dari pihak Twitter melalui teknik *crawling*. Hasil dari *crawling* data kemudian disimpan ke dalam Excel. Kemudian dilakukan *cleansing* dan *preprocessing data* untuk menjadikan sederhana dimensi dari *dataset*. Hasil dari *tweet* yang sudah diproses dan masuk kategori opini akan dipisahkan menjadi 3 kategori sentimen. Yaitu sentimen yang mempunyai polaritas positif, negatif dan netral. Hasil *tweet* yang dipisahkan tadi nantinya akan digunakan sebagai *dataset training* di mana proses pemisahannya dilakukan secara manual. Selanjutnya *dataset* tersebut disimpan dalam bentuk *file* CSV dan ARFF dan di *input* pada WEKA. Setelah seluruh proses dilakukan, barulah klasifikasi *tweet* menggunakan metode *Naïve Bayes Classifier* dimulai.

### **Preprocessing Data**

Sebelum *dataset tweet* siap digunakan maka akan terlebih dahulu dilakukan *pre-processing data* sehingga *dataset* telah bersih dan siap digunakan dalam proses selanjutnya. Pada tahap *pre-processing data*, dilakukan tahap *text mining* yang di mana melewati 9 tahap penambangan teks. Yaitu proses ekstraksi pola pada sebuah kalimat, yang di mana akan menghasilkan kalimat yang benar. Berikut adalah 9 tahap penambangan teks yang akan dilakukan [5].

1. **Normalisasi Fitur.**  
Dalam proses *Normalisasi Fitur* peneliti akan membuang komponen yang biasa terdapat pada sebuah *tweet* yaitu, *username*, URL, RT (tanda *retweet*), dan *Hashtag*.
2. **Tokenizer**  
Dalam proses *Tokenizer* peneliti akan memisahkan sebuah kalimat menjadi per kata atau *word*.
3. **Case-Folding**  
Dalam proses *Case-Folding* peneliti akan mengubah semua huruf besar menjadi huruf kecil atau *lowercase*.
4. **Clean Number**  
Dalam proses *Clean Number* peneliti akan menghapus setiap angka yang berada di depan atau di akhir sebuah kata.
5. **Convert Word**  
Dalam proses *Convert Word* peneliti akan menyempurnakan sebuah kata yang tidak baku menjadi kata baku.
6. **Convert Number**  
Dalam proses *Convert Number* peneliti akan menyempurnakan sebuah kata di mana memiliki angka di pertengahan kata yang dijadikan sebuah karakter huruf yang sebenarnya.

7. *Filtering*

Dalam proses *Filtering* peneliti akan menghapus seluruh *stop-word* (yang, dan, di, dari, dll.) yang terdapat dalam kalimat.

8. *Phrase Detection*

Tahap ini bertujuan untuk menemukan 2 kata atau lebih yang merupakan frasa kata.

9. *Stemming*

Pada proses *stemming* terhadap contoh kalimat yang ditentukan terdapat tiga kata yang memiliki kata dasar.

**Penentuan *Class Attribute***

Setelah melakukan *preprocessing* pada data *tweet* selanjutnya adalah menentukan *class attribute*. Tujuan dari *class attribute* adalah memberikan parameter kepada masyarakat terhadap objek tertentu secara akurat. *Class attribute* digunakan dalam penelitian ini ada 3 aspek, yaitu: Sentimen Positif, Sentimen Netral, dan Sentimen Negatif [3]. Hasil penentuan *Class Attribute* dapat dilihat pada Tabel 1.

**Tabel 1** Pemberian *Class Attribute* Positif, Negatif, dan Netral

| No. | <i>Tweet</i>   | Sentimen |
|-----|--|----------|
| 1   | desa maju sejak pemerintah jokowi angka nganggur turun   | Netral   |
| 2   | semoga sehat bapak presiden abdi untuk negara  | Positif  |
| 3   | presiden jokowi resmi bangkit listrik tenaga mesin gas arun daya listrik seratus delapan puluh empat mega watt untuk butuh warga lhokseumawe                             | Netral   |
| 4   | mantap sekali semoga pak jokowi kenan terima   | Positif  |
| 5   | apa menurutmu terlihat baik  | Positif  |
| 6   | prabowo tampil percaya diri sebagai bicara utama acara the world in two thousand and nineteen gala dinner singapura dua puluh tujuh bulan sebelas dua ribu delapan belas | Positif  |
| 7   | maaf pak prabowo sandiaga uno saya bisa transfer segini ya pak percaya saya terus juang doa untuk menang bapak   | Netral   |
| 8   | segar adek baju biru lupa pilih presiden baru nomor dua  | Netral   |
| 9   | empek empek kau berdebat urus negara   | Negatif  |
| 10  | biarkan prabowo sandiuno juang sendiri rapat baris indonesia adil makmur   | Positif  |

**Proses *Training Metode Naive Bayes Menggunakan WEKA***

Proses *training* ialah proses pengujian data yang telah ditentukan *Class Attribute*-nya secara manual sebagai kamus data dalam penentuan proses *testing* terhadap sentimen dengan nilai *conditional probabilities* [6]. Sehingga diperoleh hasil sentimen yang lebih akurat baik itu positif, negatif, dan netral. Dalam penelitian ini diambil 10 *tweet* yang dijadikan sebagai object *data training*. Persamaan *Naive Bayes* yang dipakai terhadap proses *training* pada WEKA ialah persamaan *Multinomial Naive Bayes*. Ada pun persamaannya sebagai berikut

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq nd} P(t_k|c) \quad (1)$$

Keterangan:

$P(t_k|c)$  : *conditional probabilities* dari kata  $t_k$ .

$P(c)$  : *prior probabilities* dari dokumen kelas  $c$ .

Untuk memperoleh nilai dari pada *prior probabilities* dinyatakan dalam persamaan :

$$P(c) = \frac{N_c}{N} \quad (2)$$

Keterangan:

$N_c$  : Jumlah dokumen *training* dalam kelas  $c$ .

$N$  : Jumlah keseluruhan dokumen *training* dari seluruh kata.

Dan untuk persamaan dari pada *conditional probabilities* ialah:

$$P(t|c) = \frac{T_{ct}}{\sum_{t' \in V} T_{ct'}} \quad (3)$$

Keterangan :

$T_{ct}$  = Jumlah kemunculan kata  $t$  dalam sebuah dokumen *training* pada kelas  $c$ .

$\frac{T_{ct}}{\sum_{t' \in V} T_{ct'}}$  = Jumlah total keseluruhan kata dalam dokumen *training* pada kelas  $c$ .

$t'$  = Jumlah total kata dalam dokumen *training*.

Untuk menghilangkan nilai nol pada sebuah dokumen, digunakannya *laplace smoothing* sebagai proses penambahan nilai 1 pada setiap nilai  $T_{ct}$  pada perhitungan *conditional probabilities* dan dinyatakan dalam persamaan:

$$P(t|c) = \frac{T_{ct} + 1}{\sum_{t' \in V} T_{ct'} + B'} \quad (4)$$

$B'$  = Total kata unik pada keseluruhan kelas dalam dokumen *training*.

Untuk memperoleh nilai probabilitas yang tinggi setiap kata digunakan *laplace smoothing* atau *add-one*, *laplace smoothing* digunakan agar nilai dari probabilitas masing-masing kata dapat memenuhi syarat yaitu tidak sama dengan 0. Jika nilai dari probabilitas kata adalah 0 maka data *training* maupun *testing* tidak akan pernah cukup untuk mewakili *frekuensi* saat terdapat kejadian langka.

**Tabel 2** Contoh Data *Training*

| No. | Tweet  | Sentimen |
|-----|--|----------|
| 1   | desa maju sejak pemerintah jokowi angka nganggur turun   | Positif  |
| 2   | semoga sehat bapak presiden abdi untuk negara  | Positif  |
| 3   | presiden jokowi resmi bangkit listrik tenaga mesin gas arun daya listrik seratus delapan puluh empat mega watt untuk butuh warga lhokseumawe                             | Netral   |
| 4   | mantap sekali semoga pak jokowi kenan terima   | Positif  |
| 5   | apa menurutmu terlihat baik  | Positif  |
| 6   | prabowo tampil percaya diri sebagai bicara utama acara the world in two thousand and nineteen gala dinner singapura dua puluh tujuh bulan sebelas dua ribu delapan belas | Positif  |
| 7   | maaf pak prabowo sandiaga uno saya bisa transfer segini ya pak percaya saya terus juang doa untuk menang bapak   | Netral   |
| 8   | segar adek baju biru lupa pilih presiden baru nomor dua  | Netral   |
| 9   | empek empek kau berdebat urus negara   | Negatif  |

**Tabel 3** Hasil Perhitungan *Prior probabilities Class Attribute* Untuk Setiap Kelas

| Atribut Kelas | P(class) |
|---------------|----------|
| Positif       | 5/9      |
| Negatif       | 1/9      |
| Netral        | 3/9      |

**Tabel 4** Hasil Perhitungan *Prior probabilities Term* Untuk Setiap Kelas

| Term       | Positif | Negatif | Netral |
|------------|---------|---------|--------|
| Desa       | 1/53    | 0/5     | 0/50   |
| Sejak      | 1/53    | 0/5     | 0/50   |
| Maju       | 1/53    | 0/5     | 0/50   |
| Pemerintah | 1/53    | 0/5     | 0/50   |
| Angka      | 1/53    | 0/5     | 0/50   |
| Nganggur   | 1/53    | 0/5     | 0/50   |
| Turun      | 1/53    | 0/5     | 0/50   |
| Mantap     | 1/53    | 0/5     | 0/50   |
| Terima     | 1/53    | 0/5     | 0/50   |
| Sehat      | 1/53    | 0/5     | 0/50   |
| Baik       | 1/53    | 0/5     | 0/50   |
| Tampil     | 1/53    | 0/5     | 0/50   |
| Percaya    | 1/53    | 0/5     | 1/50   |
| Utama      | 1/53    | 0/5     | 0/50   |

|          |      |     |      |
|----------|------|-----|------|
| Maaf     | 0/53 | 0/5 | 1/50 |
| Menang   | 0/53 | 0/5 | 1/50 |
| Doa      | 0/53 | 0/5 | 1/50 |
| Untuk    | 0/53 | 0/5 | 1/50 |
| Warga    | 0/53 | 0/5 | 1/50 |
| Berdebat | 0/53 | 1/5 | 0/50 |
| Negara   | 1/53 | 1/5 | 0/50 |

**Tabel 5** Hasil Perhitungan *Prior probabilities Term* Untuk Setiap Kelas Dalam Bentuk Desimal

| Term       | Positif | Negatif | Netral  |
|------------|---------|---------|---------|
| Desa       | 0.01242 | 0.00884 | 0.00632 |
| Sejak      | 0.01242 | 0.00884 | 0.00632 |
| Maju       | 0.01242 | 0.00884 | 0.00632 |
| Pemerintah | 0.01242 | 0.00884 | 0.00632 |
| Angka      | 0.01242 | 0.00884 | 0.00632 |
| Nganggur   | 0.01242 | 0.00884 | 0.00632 |
| Turun      | 0.01242 | 0.00884 | 0.00632 |
| Mantap     | 0.01242 | 0.00884 | 0.00632 |
| Terima     | 0.01242 | 0.00884 | 0.00632 |
| Sehat      | 0.01242 | 0.00884 | 0.00632 |
| Baik       | 0.01242 | 0.00884 | 0.00632 |
| Tampil     | 0.01242 | 0.00884 | 0.00632 |
| Percaya    | 0.01242 | 0.00884 | 0.01265 |
| Utama      | 0.01242 | 0.00884 | 0.00632 |
| Maaf       | 0.00621 | 0.00884 | 0.01265 |
| Menang     | 0.00621 | 0.00884 | 0.01265 |
| Doa        | 0.00621 | 0.00884 | 0.01265 |
| Untuk      | 0.00621 | 0.00884 | 0.01265 |
| Warga      | 0.00621 | 0.00884 | 0.01265 |
| Berdebat   | 0.00621 | 0.01769 | 0.00632 |
| Negara     | 0.01242 | 0.01769 | 0.00632 |

**Proses Testing Metode Naive Bayes Menggunakan WEKA**

Alur pada proses *testing* kurang lebih sama dengan alur *training*. Hanya yang membedakan ialah pada saat proses *testing* selesai akan dihitung nilai probabilitas akhir. Data *testing* terdapat pada Tabel 6.

**Tabel 6** Data *Testing*

| No. | Tweet  | Sentimen |
|-----|--|----------|
| 0   | biarkan prabowo sandiuno juang sendiri rapat baris indonesia adil makmur | ?        |

Perhitungan probabilitas dihitung terlebih dahulu untuk memudahkan kalkulasi selanjutnya. Untuk hasilnya terdapat pada Tabel 7. Pada sebuah *term* yang kemunculannya lebih dari satu kali, nilai kemunculan tersebut menjadi pangkat dari nilai *conditional probabilities*-nya pada Tabel 7. dengan *term frequency* (tf) masing-masing. Kemudian jumlahkan nilainya terhadap masing-masing kelas.

**Tabel 7** Hasil Nilai *Conditional probabilities*

| Term      | Tf | Positif | Negatif | Netral  |
|-----------|----|---------|---------|---------|
| Biarkan   | 1  | 0.00621 | 0.00884 | 0.00632 |
| Juang     | 2  | 0.00621 | 0.00884 | 0.01265 |
| Sendiri   | 1  | 0.00621 | 0.00884 | 0.00632 |
| Rapat     | 1  | 0.00621 | 0.00884 | 0.00632 |
| Baris     | 1  | 0.00621 | 0.00884 | 0.00632 |
| Indonesia | 1  | 0.00621 | 0.00884 | 0.00632 |
| Adil      | 1  | 0.00621 | 0.00884 | 0.00632 |
| Makmur    | 1  | 0.00621 | 0.00884 | 0.00632 |

Misalnya *term* "juang" memiliki kemunculan sebanyak 2 kali. Kemudian setiap probabilitas dari setiap kelas akan dipangkatkan dengan 2. Untuk nilai kemunculan yang sama agar proses kalkulasi tidak

membebani maka proses tersebut dilakukan satu kali. Hal ini mempercepat proses pencarian probabilitas. Pada Tabel 8 merupakan hasil seluruh pemangkatan *term* yang ditemukan.

**Tabel 8.** Hasil Perkalian Nilai Conditional Probabilities Dengan Term Frequency-nya

| Term      | Positif | Negatif | Netral  |
|-----------|---------|---------|---------|
| Biarkan   | 0.00621 | 0.00884 | 0.00632 |
| Juang     | 0.00003 | 0.00007 | 0.00016 |
| Sendiri   | 0.00621 | 0.00884 | 0.00632 |
| Rapat     | 0.00621 | 0.00884 | 0.00632 |
| Baris     | 0.00621 | 0.00884 | 0.00632 |
| Indonesia | 0.00621 | 0.00884 | 0.00632 |
| Adil      | 0.00621 | 0.00884 | 0.00632 |
| Makmur    | 0.00621 | 0.00884 | 0.00632 |
| Total     | 0.0435  | 0.06195 | 0.0444  |

Pada Tabel 8 didapatkan nilai probabilitas setiap *term*. Untuk mendapatkan kelas yang diinginkan dari dokumen adalah dengan mengalikan probabilitas dokumen dengan probabilitas kelas yang didapat dari hasil perhitungan. Kemudian diambil nilai paling besar di antara ketiga kelas tersebut [7]. Sebagai contoh *tweet* ke 10. Didapatkan nilai probabilitas dari *tweet* 10 terhadap seluruh sentimen dengan cara mengalikan nilai *prior probabilities* dengan total nilai *conditional probabilities* untuk masing-masing kelas.

*Probabilities tweet* 10 pada sentimen positif :

$$P(\text{positif} | \text{tweet 10}) = \frac{5}{9} \times 0.0435 = 0.02416$$

*Probabilities tweet* 10 pada sentimen negatif :

$$P(\text{negatif} | \text{tweet 10}) = \frac{1}{9} \times 0.0435 = 0.00483$$

*Probabilities tweet* 10 pada sentimen netral :

$$P(\text{netral} | \text{tweet 10}) = \frac{3}{9} \times 0.0435 = 0.0145$$

Dari hasil perhitungan probabilitas di atas diketahui *probabilities tweet* 10 terdapat pada sentimen positif karena memiliki nilai yang paling besar. Oleh sebab itu *tweet* 10 terklasifikasi pada sentimen positif.

### **Performing Measure**

Langkah yang terakhir dari penelitian ini merupakan proses analisis dari sistem yang kita buat untuk mengukur akurasi dari sistem yang kita rancang sendiri. Pada tugas akhir kali ini proses pengukuran performansi diukur dari *Precision, Recall, F-Measure* dan Akurasi.

#### **1. Confusion Matrix**

Merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi pada dasarnya *confusion matrix* itu sendiri mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Tabel 9 adalah *Confusion Matrix* dari hasil klasifikasi *Naïve Bayes* yang sudah terklasifikasi.

**Tabel 9** *Confusion Matrix.*

|        |         | Predicted |         |        |
|--------|---------|-----------|---------|--------|
|        |         | Positif   | Negatif | Netral |
| Actual | Positif | TP        | FP      | FP     |
|        | Negatif | FN        | TN      | FN     |
|        | Netral  | FNe       | FNe     | TNe    |

2. *Precision*

*Precision* adalah jumlah data yang benar diklasifikasikan dengan sebuah kelas dibagi dengan jumlah total data yang telah diklasifikasikan ke dalam kelas tersebut dengan rumus sebagai berikut. Nilai *precision* digunakan untuk mengukur ketepatan rekomendasi yang dihasilkan oleh sistem.

$$Precision = \frac{TP}{(TP+FP)} \quad (5)$$

Di mana:

TP adalah *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem  
 FP adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem

3. *Recall*

*Recall* adalah jumlah data yang benar diklasifikasikan dalam sebuah kelas dibagi dengan jumlah total data dalam kelas tersebut.

$$Recall = \frac{TP}{(TP+FN)} \quad (6)$$

Di mana:

TP adalah *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.  
 FN adalah *False Negative*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.

4. *F-Measure*

*F-Measure* digunakan untuk mengevaluasi rata-rata nilai *Precision* dan *Recall* hasil klasifikasi.

$$F - Measure = 2 x \frac{Precision \times Recall}{Precision+Recall} \quad (7)$$

Di mana perhitungannya adalah 2 dikali dengan *Precision* dikali dengan hasil *Recall* dibagi jumlah *Precision* ditambah *Recall*

5. *Akurasi*

*Akurasi* merupakan parameter evaluasi terhadap sistem yang dibangun dalam penelitian tugas akhir ini. Berikut adalah rumus akurasi.

$$Akurasi = TP + TN / (TP + FP + TN + FN) \quad (8)$$

Di mana:

TP adalah *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.  
 TN adalah *True Negative*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.  
 FN adalah *False Negative*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.  
 FP adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem

### 3. Hasil

Setelah melakukan beberapa penelitian untuk menganalisa *tweet* menggunakan algoritma *Naïve Bayesian Classifier*, maka didapatkan hasil Analisa sebagai berikut:

| === Confusion Matrix === |    |    |                   | Correctly Classified Instances   | 628    | 77.6267 % |
|--------------------------|----|----|-------------------|----------------------------------|--------|-----------|
|                          |    |    |                   | Incorrectly Classified Instances | 181    | 22.3733 % |
| a                        | b  | c  | <-- classified as | Precision                        | Recall | F-Measure |
| 572                      | 0  | 0  | a = Positif       | 0.761                            | 1.000  | 0.864     |
| 168                      | 30 | 1  | b = Netral        | 1.000                            | 0.151  | 0.262     |
| 12                       | 0  | 26 | c = Negatif       | 0.963                            | 0.684  | 0.800     |
|                          |    |    |                   | 0.829                            | 0.776  | 0.713     |

**Gambar 2** Detail Klasifikasi Naive Bayesian Classification

Pada Gambar 2 dapat dianalisa bahwa *Correctly Classified Instances* merupakan suatu parameter dari tingkat akurasi yang benar (77,62 % = 628 data *tweet*) dan tingkat dari kesalahan atau disebut dengan *Incorrectly Classified Instances* (22,37% = 181 data *tweet*) yang bisa kita analisa dengan:



$$\frac{\text{Jumlah status yang diklasifikasi benar dan memang benar}}{\text{Jumlah status (tweet) secara keseluruhan}} \times 100\%$$

$$\frac{628 \text{ tweet}}{809 \text{ tweet}} \times 100\% = 0.7762 = 77,62\%$$

Sedangkan untuk *Incorrectly Classified Instances* berikut dengan perhitungannya

$$\frac{\text{Jumlah status yang diklasifikasi salah dan memang salah}}{\text{Jumlah status (tweet) secara keseluruhan}} \times 100\%$$

$$\frac{181}{809} \times 100\% = 0.2237 = 22.37\%$$

Penjelasan *Confusion Matrix*:

1. Baris pertama "572 - 0 - 0" ada 572 *tweet* yang diprediksi sebagai sentimen positif dan memang benar merupakan sentimen positif.
2. Baris kedua "168 - 30 - 1" ada 199 *tweet* yang diprediksi seluruhnya sebagai sentimen netral tetapi 129 *tweet* di antaranya terklasifikasi sebagai sentimen positif, 30 *tweet* merupakan sentimen yang benar netral, dan 1 terklasifikasi sebagai sentimen negatif.
3. Baris ketiga "12 - 0 - 26" ada 38 *tweet* yang diprediksi sebagai sentimen negatif tetapi 12 *tweet* di antaranya terklasifikasi sebagai sentimen positif dan 26 di antaranya memang benar merupakan sentimen negatif.

| No. 1: Tweet<br>Nominal | 2: prediction margin<br>Numeric | 3: predicted Sentimen<br>Nominal | 4: Sentimen<br>Nominal |
|-------------------------|---------------------------------|----------------------------------|------------------------|
| ... pak j...            | 0.476966                        | Positif                          | Positif                |
| ... suka ...            | 0.596605                        | Positif                          | Positif                |
| ... duku...             | 0.476966                        | Positif                          | Positif                |
| ... prest...            | 0.6717                          | Positif                          | Positif                |
| ... jalan ...           | 0.476966                        | Positif                          | Positif                |
| ... yang j...           | 0.723225                        | Positif                          | Positif                |
| ... yang j...           | 0.723225                        | Positif                          | Positif                |
| ... harga...            | 0.476966                        | Positif                          | Positif                |
| ... jokow...            | 0.476966                        | Positif                          | Positif                |
| ... jokow...            | 0.596605                        | Positif                          | Positif                |
| ... jokow...            | 0.476966                        | Positif                          | Positif                |
| ... pak j...            | 0.476966                        | Positif                          | Positif                |
| ... perny...            | 0.476966                        | Positif                          | Positif                |
| ... maki...             | 0.476966                        | Positif                          | Positif                |
| ... prest...            | 0.6717                          | Positif                          | Positif                |
| ... jelas ...           | 0.476966                        | Positif                          | Positif                |
| ... semo...             | 0.476966                        | Positif                          | Positif                |
| ... jokow...            | 0.476966                        | Positif                          | Positif                |
| ... untuk...            | 0.476966                        | Positif                          | Positif                |
| ... namu...             | 0.476966                        | Positif                          | Positif                |
| ... setuju...           | 0.476966                        | Positif                          | Positif                |
| ... hebat...            | 0.596605                        | Positif                          | Positif                |
| ... mirip ...           | 0.596605                        | Positif                          | Positif                |
| ... masy...             | 0.476966                        | Positif                          | Positif                |
| ... jokow...            | 0.596605                        | Positif                          | Positif                |
| ... alha...             | 0.596605                        | Positif                          | Positif                |
| ... kalau...            | 0.596605                        | Positif                          | Positif                |
| ... hebat...            | 0.596605                        | Positif                          | Positif                |
| ... jokow...            | 0.476966                        | Positif                          | Positif                |
| ... jokow...            | 0.596605                        | Positif                          | Positif                |
| ... alha...             | 0.476966                        | Positif                          | Positif                |
| ... sekto...            | 0.476966                        | Positif                          | Positif                |
| ... karen...            | 0.476966                        | Positif                          | Positif                |
| ... pak j...            | 0.476966                        | Positif                          | Positif                |
| ... prest...            | 0.6717                          | Positif                          | Positif                |
| ... jokow...            | 0.596605                        | Positif                          | Positif                |
| ... yang j...           | 0.723225                        | Positif                          | Positif                |

**Gambar 3** Detail Nilai *Conditional probabilities* Klasifikasi *Naive Bayesian Classification*

Maka jumlah data *precision*, *recall*, *f-measure*, akurasi *tweet* terhadap Bapak Joko Widodo adalah:

1. *Precision*

$$Precision = \frac{572}{(572 + 168 + 12)} = 0,76$$

2. *Recall*

$$Recall = \frac{572}{(572 + 0 + 0)} = 1,00$$

3. *F-Measure*

$$F-Measure = 2 \times \frac{1 \times 0,76}{1 + 0,76} = 0.86$$

4. *Akurasi*

$$Akurasi = \frac{628}{808} = 0,77$$

Dan persentase jumlah data sentimen *tweet* terhadap Bapak Joko Widodo adalah sebagai berikut:

$$Hasil\ Prediksi = \frac{Jumlah\ tweet\ (sentimen)}{Jumlah\ keseluruhan\ tweet} \times 100\%$$

A. Sentimen Positif terhadap Bapak Joko Widodo.

$$Hasil = \frac{753}{809} \times 100\% = 0.9307 = 93\%$$

B. Sentimen Netral terhadap Bapak Joko Widodo.

$$Hasil = \frac{30}{809} \times 100\% = 0.0370 = 4\%$$

C. Sentimen Negatif terhadap Bapak Joko Widodo.

$$Hasil = \frac{27}{809} \times 100\% = 0.0321 = 3\%$$

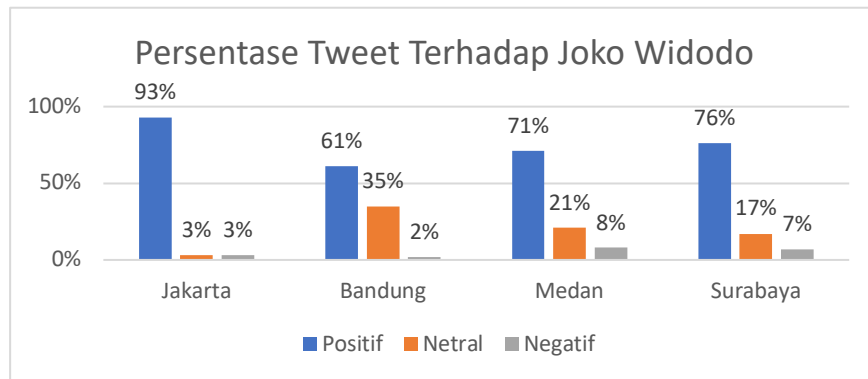
Berdasarkan data yang diperoleh pada setiap *geolocation* yang ditentukan pada Tabel 11 dan 22 dari pada *hashtag* **#jokowi2periode** dan **#prabowosandi** diperoleh hasil persentase diagram terhadap dua tokoh tersebut.

**Tabel 11.** Jumlah *Tweet* Mengenai #jokowi2periode Dari Jakarta, Bandung, Surabaya, dan Medan

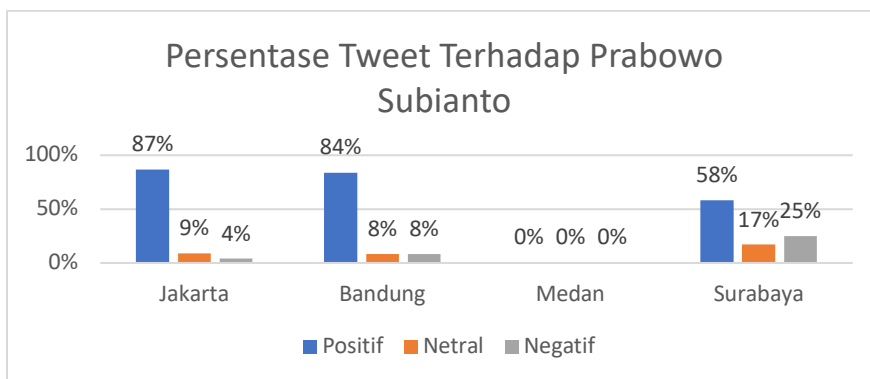
| Kota         | Jakarta | Bandung | Surabaya | Medan | Total |
|--------------|---------|---------|----------|-------|-------|
| <i>Tweet</i> | 809     | 62      | 28       | 39    | 938   |

**Tabel 12.** Jumlah *Tweet* Mengenai #prabowosandi Dari Jakarta, Bandung, Surabaya, dan Medan

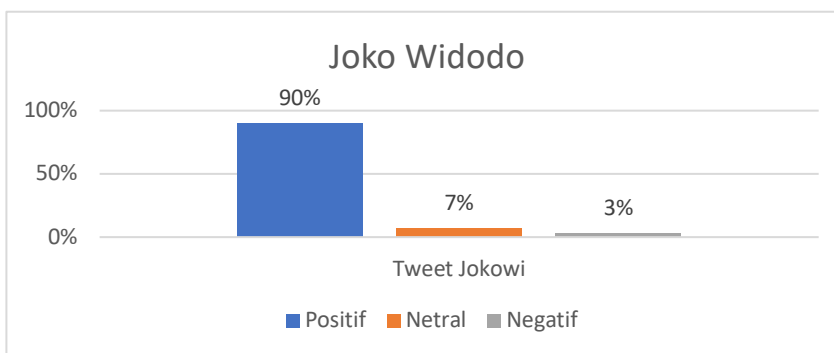
| Kota         | Jakarta | Bandung | Surabaya | Medan | Total |
|--------------|---------|---------|----------|-------|-------|
| <i>Tweet</i> | 55      | 13      | 12       | 0     | 80    |



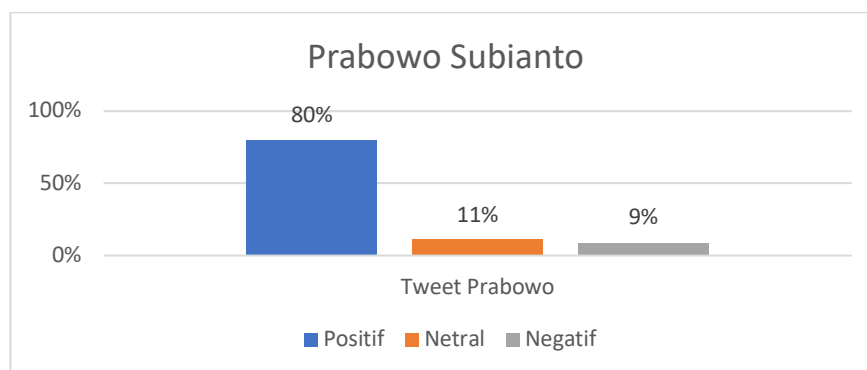
**Gambar 4** Persentase Jumlah *Tweet* Mengenai #jokowi2periode Dari Jakarta, Bandung, Medan, dan Surabaya



**Gambar 5** Persentase Jumlah *Tweet* Mengenai #prabowosandi Dari Jakarta, Bandung, Medan, dan Surabaya



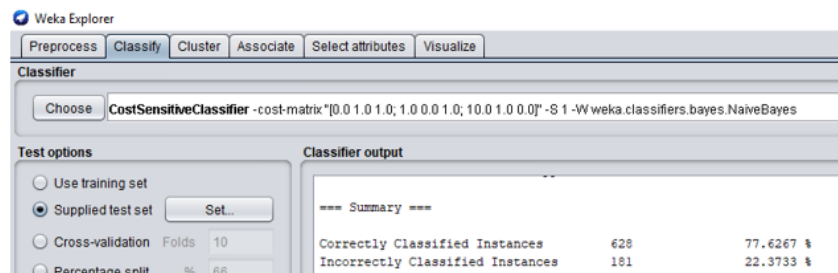
**Gambar 6** Persentase Jumlah Seluruh *Tweet* Mengenai #jokowi2periode Dari Jakarta, Bandung, Medan, dan Surabaya



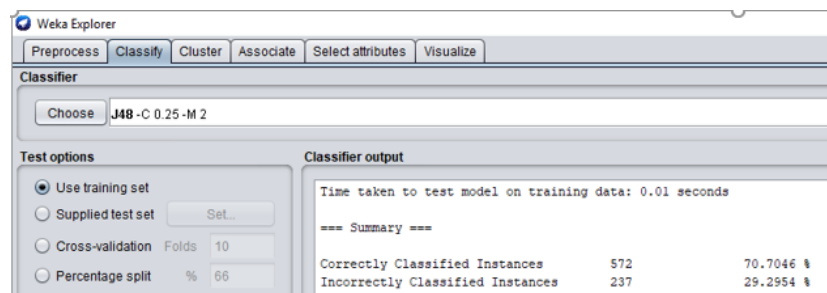
**Gambar 7** Persentase Jumlah Seluruh *Tweet* Mengenai #prabowosandi Dari Jakarta, Bandung, Medan, dan Surabaya

### Metode Pemanding

Pada penelitian ini, peneliti membuat metode pemanding sebagai acuan yang lebih akurat mengapa harus memakai metode *Naïve Bayes*. Ada pun metode pemanding yang peneliti tetapkan adalah metode *Tree.J48*. Hasil dari perbandingan ini dapat dilihat dari tingkat akurasi yang dihasilkan. Ada pun hasil akurasi dapat dilihat pada Gambar 8.



**Gambar 8** Hasil Pengujian *Data Training* Dengan Metode *Naïve Bayes*.



**Gambar 9** Hasil Pengujian *Data Training* Dengan Metode *Tree.J48*.

Dari hasil akurasi kedua metode pada Gambar 8 dan Gambar 9 dapat disimpulkan tingkat akurasi yang paling tinggi terdapat pada metode *Naïve Bayes* dengan persentase sebesar 77,62% sedangkan *Tree.J48* sebesar 70,70%. Oleh karena itu peneliti menggunakan metode *Naïve Bayes* sebagai metode klasifikasi data pada penelitian ini.

#### 4. Kesimpulan

Konfigurasi algoritma *Naïve Bayesian Classifier* terhadap pra-pemilihan presiden tahun 2019 dimulai dari pengumpulan data manual melalui *Twitter*. Lalu dilakukannya *pre-processing* terhadap data sampai diklasifikasikannya pada aplikasi weka dengan perhitungan *Naïve Bayesian Classifier* terhadap probabilitas kata. Dengan memperoleh hasil probabilitas tertinggi, sehingga dapat mengategorikan sebuah kalimat menjadi sebuah sentimen yang sebenarnya.

Perolehan sentimen *tweet* berdasarkan wilayah didapat melalui titik koordinat geografis bumi yang terhubung pada aplikasi *Twitter* melalui *API Geolocation* yang terdiri dari garis lintang dan garis bujur pada *google maps* yang telah ditetapkan oleh pengguna.

Perbandingan klasifikasi algoritma *Naïve Bayesian Classifier* dengan algoritma *Tree.J48* memiliki tingkat akurasi yang cukup berbeda. Di mana klasifikasi algoritma *Naïve Bayesian Classifier* lebih tinggi dengan tingkat akurasi 77,62% dibandingkan dengan algoritma *Tree.J48* yang hanya 70%.

#### Saran

Penulis menyarankan pengembangan penelitian lebih lanjut ialah melakukan analisis sebuah daerah secara otomatis melalui sentimen yang ada. Dengan kemungkinan pengambilan sumber data dari beberapa alternatif media sosial lainnya.

## 5. Referensi

- [1] G. A. Buntoro, A. E. Permanasari, and T. Bharata, "Sentimen Analysis Candidates of Indonesian Presiden 2014 with Five Class Attribute," *Int. J. Comput. Appl.*, vol. 136, pp. 23–29, 2016.
- [2] S. Mujilawati, "Pre-Processing Text Mining Pada Data Twitter," *Semin. Nas. Teknol. Inf. Dan Komun. 2016*, pp. 49–56, 2016.
- [3] A. F. Hidayatullah and A. Azhari, "Analisis Sentimen dan Klasifikasi Kategori terhadap Tokoh Publik pada Twitter," *Semin. Nas. Inform. 2014*, pp. 115–122, 2014.
- [4] R. McCue, *A Comparison of the Accuracy of Support Vector Machine and Naïve Bayes Algorithms In Spam Classification*. Santa Cruz: University of California, 2009.
- [5] I. Sunni and D. H. Widyantoro, "Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik," *J. Sarj. Inst. Teknol. Bdg. Bid. Tek. Elektro Dan Inform.*, vol. 1, no. 2, pp. 200–206, Jul. 2012.
- [6] J. Ling, T. B. Oka, and I. P. E. N. Kencana, "ANALISIS SENTIMEN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DENGAN SELEKSI FITUR CHI SQUARE," *E-J. Mat.*, vol. 3, pp. 92–99, Aug. 2014.
- [7] R. A. Simanjuntak, D. Gunawan, and Amalia, "Analisis Sentimen pada Layanan Gojek Indonesia Menggunakan Multinomial Naive Bayes," Universitas Sumatera Utara, Sumatera, Indonesia, 2018.