

# **ANALISIS SENTIMEN TOKOH PUBLIK MENGGUNAKAN METODE NAÏVE BAYESIAN CLASSIFICATION PADA APLIKASI TWITTER**

**Yusran Timur Samuel, Kevin Jeremy Manurip dan Heyma Sitorus  
Fakultas Teknologi Informasi, Universitas Advent Indonesia**

## **ABSTRAK**

Media sosial telah menyediakan beragam informasi khususnya konten-konten yang bersifat subjektif atau merefleksikan opini dari orang yang menulis. Dewasa ini semakin banyak orang menyatakan pendapat atau opini mereka terhadap tokoh-tokoh masyarakat seperti pemimpin wilayah, pejabat, orang berpengaruh, dan sebagainya. Media sosial telah menyediakan beragam informasi khususnya konten-konten yang bersifat subjektif atau merefleksikan opini dari orang yang menulis.

Dewasa ini semakin banyak orang menyatakan pendapat atau opini mereka terhadap tokoh-tokoh masyarakat seperti pemimpin wilayah, pejabat, orang berpengaruh, dan sebagainya. Analisis sentimen pada aplikasi sosial media Twitter terdapat kelemahan dalam kata-kata yang terdapat pada kalimat yang diunggah oleh pengguna aplikasi tersebut. Dalam hal ini objek penelitian dilakukan kepada Ridwan Kamil dengan Sentimen Analisis dari rakyat. Berdasarkan penelitian yang dilakukan, maka didapatkan kesimpulan bahwa dari 59 data yang di training memiliki akurasi sebesar 81.3559 %, dan hasil yang didapat dari data testing adalah: 1. Terdapat 3 data yang benar-benar diklasifikasikan memiliki sentiment netral, 2. Terdapat 7 data yang diklasifikasikan benar-benar memiliki sentiment positif, 3. Dan terdapt 2 data yang benar-benar memiliki sentiment negatif.

**Kata-kata kunci:** Analisis Sentimen, Naïve bayes, Naïve Bayesian Classification, Twitter

## ***SENTIMENT ANALYSIS OF PUBLIC FIGURE USING NAÏVE BAYESIAN CLASSIFICATION ON TWITTER APPLICATION***

### **ABSTRACT**

*Social media has provided a variety of information, especially content that is subjective or reflects the opinions of people who write. Today more and more people express their opinions or opinions on community leaders such as regional leaders, officials, influential people, and so on. Social media has provided a variety of information, especially content that is subjective or reflects the opinions of people who write.*

*Today more and more people express their opinions or opinions on community leaders such as regional leaders, officials, influential people, and so on. Sentiment analysis on the Twitter social media application there are weaknesses in the words contained in the sentence uploaded by the application user. In this case the object of the research was carried out to Ridwan Kamil with Sentiment Analysis from the people. Based on the research conducted, it was concluded that 59 training data had an accuracy of 81.3559%, and the results obtained from testing data were: 1. There were 3 data that were truly classified as having neutral sentiments, 2. There were 7 data classified really have positive sentiment, 3. And there are 2 data that really have negative sentiment.*

## Pendahuluan

Media sosial telah menyediakan beragam informasi khususnya konten-konten yang bersifat subjektif atau merefleksikan opini dari orang yang menulis. Hal ini dimanfaatkan oleh para peneliti untuk mengetahui pendapat orang mengenai siapa orang-orang tersebut. Pengguna internet menggunakan review terhadap subjek yang diinginkan secara online dan dari hasil review tersebut didapat bahwa pembaca dapat terpengaruh dalam menentukan pendapat serta opini mereka. Analisis sentimen sangat berguna dalam pemantauan media sosial karena memungkinkan kita untuk memperoleh gambaran tentang opini publik yang luas dari suatu topik-topik tertentu.

Dalam penelitian kali ini, penulis menggunakan metode Support Vector Machine (SVM). Analisis sentimen pada aplikasi sosial media Twitter terdapat kelemahan dalam kata-kata yang terdapat pada kalimat yang diunggah oleh pengguna aplikasi tersebut. Aplikasi Twitter hanya dapat mengunggah karakter sebanyak 140 karakter, hal ini yang menyebabkan user menggunakan perkataan serta kalmia yang disingkat atau dibuat inisial serta bisa mengarah kepada ejaan yang salah. Berdasarkan latar belakang masalah di atas maka identifikasi masalah adalah bagaimana menganalisis sentimen sebuah tweet pada Twitter yang dapat menentukan pendapat atau opini terhadap suatu objek secara otomatis.

## Landasan Teori

Secara luas pemilihan umum merupakan suatu hal yang demokratis. Proses yang dilakukan adalah memilih orang yang akan mengisi jabatan-jabatan dalam suatu bidang politik tertentu. Jabatan-jabatan tersebut beraneka ragam, dimulai dari yang tertinggi yaitu presiden, gubernur, walikota, bupati, camat, lurah, kepala desa, RT, dan RW.

Menjelang pemilihan umum biasanya para calon-calon pemimpin akan diberikan waktu untuk mempromosikan dirinya dalam sebuah kampanye. Kegiatan ini memiliki sifat dan tujuan yang persuasif, yaitu mengajak orang-orang untuk memilih para calon untuk menjadi pemimpin dalam suatu wilayah tertentu. Hubungan antar politik juga harus kuat sehingga dapat menyokong kegiatan selama kampanye berlangsung.

Menurut (H. Hasbullah, H Nasution, M. Saleh, 2015). Pemilihan umumnya mempunyai esensi sebagai sarana demokrasi untuk membentuk sistem kekuasaan negara yang pada dasarnya lahir dari bawah menurut kehendak rakyat sehingga terbentuk kekuasaan negara yang benar-benar memancarkan ke bawah sebagai sesuatu kewibawaan sesuai dengan keinginan rakyat, oleh rakyat, menurut sistem permusyawaratan dan perwakilan. Pada hakekatnya, pemilu merupakan pengakuan dan perwujudan dari hak-hak politik rakyat dan sekaligus merupakan pendelegasian hak-hak tersebut oleh rakyat kepada wakil-wakilnya untuk menjalankan pemerintahan.

## Data Mining

Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data. Tahapan dari proses *Knowledge Discovery in Database*(KDD) adalah : 1. Selection 2. Pre-Processing / Cleaning. 3. Transformation 4. Data Mining 5. Interpretation / Evaluation.

## Metode Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Model itu sendiri bisa berupa aturan "jika-maka", berupa pohon keputusan, atau formula matematis.

## **Text Mining**

*Text mining* (penambangan Teks) adalah bidang baru yang cukup menarik perhatian di bidang dunia komputer yang diimplementasikan untuk memecahkan krisis informasi yang berlebihan dengan menggunakan penggabungan teknik dari data mining, pembelajaran mesin, pengolahan bahasa alami, pencarian informasi, dan manajemen pengetahuan.

## **Pre-processing**

Penelitian yang dilakukan oleh dua orang mahasiswa ITB, Ismail Sunni dan Dwi Hendratmo Widyanto (Sunni & Widyantoro, 2012) dalam journal mereka mengenai Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik, ada 9 tahap praproses data pada text mining, yaitu:

1. *Tokenizer*  
Tokenizer merupakan proses untuk memisah-misahkan kata atau word. Proses tokenizer dalam Twitter memiliki perbedaan dengan proses tokenizer pada teks lain. Hal ini dikarenakan adanya emoticon yang sering digunakan oleh pengguna Twitter dalam mengungkapkan perasaan.
2. *Normalisasi Fitur*  
Adanya beberapa komponen khas yang biasa ada di tweet yakni, username, URL, "RT" (tanda retweet), dan Hashtag. Karena username, URL, dan "RT" tidak memiliki komponen di atas, maka akan dibuang.
3. *Casefolding*  
Model kesalahan orthographic model, akan dikoreksi dengan cara mengubah semua huruf menjadi lowercase atau huruf kecil.
4. *Clean Number*  
Kemunculan angka di depan dan akhir kata memiliki pengaruh yang tidak signifikan dari nilai sentimen suatu tweet.
5. *Conver Number*  
Kemunculan angka ditengah-tengah kata perlu diubah menjadi karakter serupa.
6. *Conversation Emoticon*  
Emoticon merupakan salah satu cara pengungkapan perasaan secara tekstual. Hal ini tentu akan membantu dalam menentukan sentimen suatu tweet. Setiap emoticon akan dikonversi ke dalam string. Namun, tidak semua emoticon akan dikonversi dan diimplementasikan, karena tidak semua emoticon sering digunakan oleh pengguna Twitter.
7. *Stop Word Removal*  
Proses ini akan menghilangkan kata-kata yang sering muncul namun tidak memiliki pengaruh apapun dalam ekstraksi sentimen suatu tweet.

## **Algoritma Naive Bayes**

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris. *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Persamaan dari teorema *Bayes* adalah :

$$P(s|x) = \underset{x \in X}{\operatorname{argmax}} \frac{P(y|x)P(x)}{P(X)}$$

$$= \frac{\text{argmax}}{x \in X} P((y|x) P(X)$$

## Metodologi Penelitian

### Crawling Data

Dari pembahasan bab sebelumnya telah dijabarkan tentang teori-teori mengenai algoritma-algoritma yang mengklasifikasikan tweet secara lengkap dan saksama. Dalam penelitian kali ini terdapat beberapa proses yang diperlukan untuk menganalisa suatu data tweet dalam Twitter yang akan di Analisa. Crawling data yang didapat dari tweet mengenai bapak Basuki Tjahaya Purnama dan bapak Anies Baswedan tidak dilakukan dengan cara manual, namun menggunakan aplikasi pemrograman dengan mengintegrasikan console yang bernama Python.

Media yang digunakan sebagai pengumpulan data untuk penelitian ini penulis menggunakan Twitter API (Application Programming Interface). Aplikasi ini sudah sangat mendukung para pengembang/developer sebelumnya untuk membantu serta memudahkan dalam mengumpulkan data-data di Twitter.

```
## Access Key ##
# consumer_key      = uEPDQxDp30A28IKms3AP4Khr
# consumer_secret   = xNjyndpeaeX9YZw2q0RK5hpD6jGmozgWZu1NNPZLk46LVU4K3S
# access_token      = 915863860896161792-m3B9fp2PkM0r707Jw6BR2Pe2BBbzom
# access_token_secret = CBc1k1iR3J9SbwZW0PJ5YPPKoCqzW0WvwKGMLFmMpz2eW
=====
# ** NOTE ** #
# --> Only use for Twitter Developer to Crawling data #
```

### N-Gram

N-Gram merupakan kesalahan ejaan yang sering dilakukan pada saat melakukan penulisan pada kata. Hal ini dapat mengakibatkan salah pengertian terhadap orang-orang yang membacanya. Maka dapat disimpulkan bahwa kesalahan ejaan dapat disebabkan oleh 4 faktor dibawah ini:

1. *Subtitusi*, yaitu kesalahan yang diakibatkan penggantian satu huruf pada kata,
2. *Insterion*, yaitu kesalah yang diakibatkan terjadinya penyisipan atau penambahan satu huruf pada kata,
3. *Deletion*, adalah kesalahan yang diakibatkan hilangnya satu huruf pada kata yang dituliskan,
4. *Transposition*, yaitu kesalahan penukaran posisi dua huruf berdekatan.

### Data Knowledge

Data *Knowledge* merupakan data hasil dari proses *training* yang telah dilakukan, dan hasil yang telah didapatkan dari proses *training* ini akan digunakan sebagai data acuan untuk proses *testing*.

## Hasil Analisis

### Pengumpulan Data dan Pengolahan

Dalam penelitian ini, ada yang dikumpulkan adalah data twitter, dan jumlah data yang dikelolah adalah sebanyak 59 data twitter. Dan data yang digunakan untuk proses *testing* adalah sebanyak 30% dari data yang digunakan untuk proses *training*.

### Pengujian data

Pengujian yang dilakukan penulis untuk mengetahui sentiment adalah dengan menggunakan aplikasi WEKA. berikut ini merupakan contoh data yang belum di *preprocessing* dan data yang sudah di *text preprocessing*.

<i>preprocessing</i>	<i>Text preprocessing</i>
Kang Emil Buka Peluang Bakal Jadi Kader Golkar	bapak ridwan kamil memiliki peluang untuk menjadi kader golkar
Siapa Wakil @ridwankamil di Pilgub Jawa Barat 2018?	siapa wakil ridwan kamil di pemilihan gubernur jawa barat 2018
Tunjukkan SK ke Ridwan Kamil, Idrus Yakin Dedi Mulyadi Setia	Tunjukkan surat keputusan ke ridwan kamil idrus yakin dedi mulyadi setia
Siapakah yg lebih dipercaya melestarikan serta menjaga adat dan budaya sunda?	Siapakah yang lebih dipercaya melestarikan serta menjaga adat dan budaya sunda
Tindakan pemerintah nya mana? Apa cuma bilang bencana taunan?	Tindakan pemerintahnya mana apa cuma bilang bencana tahunan
Energi terbesar kita adalah "keyakinan"	Energi terbesar kita adalah keyakinan

**Tabel 4.1** *Text preprocessing*

### Pengujian data *Training*

Data yang *training* adalah sebanyak 59 data.

```

Time taken to test model on training data: 0.28 seconds

=== Summary ===

Correctly Classified Instances      48          81.3559 %
Incorrectly Classified Instances    11          18.6441 %
Kappa statistic                    0.6402
Mean absolute error                 0.3009
Root mean squared error             0.335
Relative absolute error             76.2915 %
Root relative squared error         75.6658 %
Total Number of Instances          59

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                1.000   0.423   0.750     1.000   0.857     0.658   1.000    1.000    positif
                1.000   0.000   1.000     1.000   1.000     1.000   1.000    1.000    netral
                0.000   0.000   0.000     0.000   0.000     0.000   1.000    1.000    negatif
Weighted Avg.   0.814   0.237   0.674     0.814   0.734     0.622   1.000    1.000

=== Confusion Matrix ===

 a  b  c  <-- classified as
33  0  0 | a = positif
 0 15  0 | b = netral
11  0  0 | c = negatif
    
```

**Gambar 4.1** Hasil *Training*

Keterangan **Gambar 4.1:**

1. Untuk nilai *correctly classified instance* adalah sebesar 81.3559 %, dengan jumlah 48 data yang diklasifikasikan benar dari 59 data *training*,
2. *Incorrectly classified instance* adalah sebesar 18.6441 %, dengan jumlah 11 data yang diklasifikasikan benar dari 59 data.

```

=== Confusion Matrix ===

 a  b  c  <-- classified as
33  0  0 | a = positif
 0 15  0 | b = netral
11  0  0 | c = negatif
    
```

**Gambar 4.2** *Confusion Matrix*

Keterangan **Gambar 4.2**

1. Pada baris pertama terdapat 33 data yang diklasifikasikan benar "positif"
2. Baris kedua, terdapat 15 data yang diklasifikasikan banar "Netral"
3. Baris ketiga, terdapat 0 data yang diklasifikasikan bersentimen *negative*, dan terdapat 11 data yang seharusnya diklasifikasikan bersentimen *negative*, namun diklasifikasikan bersentimen "positif".

### Pengujian data *Testing*

Pada Proses ini, data yang digunakan adalah 20% dari data *training*, yaitu sebanyak 12 data yang diambil secara acak.

Sentimen	Frekuensi	Total data
----------	-----------	------------

Negatif	2	12
Positif	7	12
Netral	3	12

```

Time taken to build model: 0 seconds

=== Predictions on training set ===

inst#   actual   predicted error prediction
  1 1:'netral ' 1:netral    0.456
  2 1:'netral ' 1:netral    0.456
  3 2:positif 2:positif    0.636
  4 2:positif 2:positif    0.636
  5 1:'netral ' 1:netral    0.456
  6 3:negatif 3:negatif    0.384
  7 2:positif 2:positif    0.636
  8 2:positif 2:positif    0.636
  9 3:negatif 3:negatif    0.384
 10 2:positif 2:positif    0.636
 11 2:positif 2:positif    0.636
 12 2:positif 2:positif    0.636
    
```

**Gambar 4.4** Hasil Prediksi data *testing*

```

=== Confusion Matrix ===

a b c  <-- classified as
3 0 0 | a = netral
0 7 0 | b = positif
0 0 2 | c = negatif
    
```

**Gambar 4.5** *Confusion Matrix testing*

### Hasil Penelitian

Berikut ini merupakan hasil penelitian yang telah dilakukan menggunakan aplikasi WEKA.

1. Terdapat 3 data yang benar-benar diklasifikasikan memiliki sentiment netral,
2. Terdapat 7 data yang diklasifikasikan benar-benar memiliki sentiment positif,
3. Dan terdapt 2 data yang benar-benar memiliki sentiment negatif.

Maka dari keterangan diatas, dapat disimpulkan bahwa tingkat kebenaran dari proses yang dilakukan WEKA adalah 100%.

### Kesimpulan dan Saran

Berdasarkan penelitian yang dilakukan, maka didapatkan kesimpulan bahwa dari 59 data yang di *training* memiliki akurasi sebesar 81.3559 %, dan hasil yang didapat dari data *testing* adalah:

1. Terdapat 3 data yang benar-benar diklasifikasikan memiliki sentiment netral,
2. Terdapat 7 data yang diklasifikasikan benar-benar memiliki sentiment positif,
3. Dan terdapt 2 data yang benar-benar memiliki sentiment negatif.

### **Saran**

1. Untuk penelitian selanjutnya disarankan untuk membandingkan dengan metode *Decisiom tree*-J48, K-Star, dan yang lainnya.
2. Membuat aplikasi, sehingga dapat menentukan sentiment dengan lebih mudah.

### **Daftar Pustaka**

- H. Hasbullah, H Nasution, M. Saleh, L. P. S. ishlahiyah B. (2015). Laporan riset partisipasi pemilih pada pemilihan umum.
- Sunni, I., & Widyantoro, D. H. (2012). Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik. *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro Dan Informatika*, 1(2), 200–206.