

# Analisis Sentimen Terhadap Pemerintahan Ridwan Kamil Sebagai Gubernur Jawa Barat Menggunakan Algoritma Naïve Bayes

## *Sentiment Analysis Of Ridwan Kamil Government As West Java Governor Using Naïve Bayes Algorithm*

U. Darmanto Soer<sup>1</sup>, Sutrisno<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

<sup>1</sup>darmantosoer@pelitabangsa.ac.id\*, <sup>2</sup>sutrisno94@gmail.com

### **Abstract**

*The use of social media in the era of globalization is very necessary for some circles including the regional leader, in the period of his tenure, Ridwan Kamil received various inputs and criticisms, and in this case the authors conducted research to analyze public sentiments towards the elected governor. And in this study the authors use the Naïve Bayes algorithm to classify sentiments and look for the preference values because the algorithm has a pretty good accuracy. From the results of tests conducted using cross validation techniques and accuracy measurements using confusion matrix with 10 times the best accuracy testing obtained was 84.38% and the positive response obtained from the calculation of preference value was 49%. Thus it can be concluded that the Naïve Bayes algorithm can be used to classify quite well and be able to measure the community's response to regional leaders.*

**Keywords:** *Sentiment Analysis, Twitter, Naïve Bayes Classifier, Cross Validation, Preference Value*

### **Abstrak**

Penggunaan sosial media di era globalisasi sangat diperlukan bagi sebagian kalangan tidak terkecuali pemimpin daerah, dalam setaun masa jabatannya ridwan kamil mendapatkan berbagai pujian masukan maupun kritikan dan dalam hal ini penulis melakukan penelitian untuk menganalisa sentimen masyarakat terhadap gubernur terpilih. Dan pada penelitian ini penulis menggunakan algoritma Naïve Bayes untuk melakukan klasifikasi sentimen dan mencari nilai preference value dikarenakan algoritma tersebut memiliki akurasi yang cukup baik. Dari hasil pengujian yang dilakukan menggunakan teknik cross validation dan pengukuran akurasi menggunakan confusion matrix dengan dilakukan 10 kali pengujian akurasi terbaik yang diperoleh adalah 84,38 % dan respon positif yang didapatkan dari hasil penghitungan preference value adalah 49%. Dengan demikian dapat disimpulkan bahwa algoritma Naïve Bayes dapat digunakan untuk melakukan klasifikasi dengan cukup baik dan dapat mengukur respon masyarakat terhadap pemimpin daerah.

**Kata kunci:** *Analisa Sentimen, Twitter, Naïve Bayes Classifier, Cross Validation, Preference Value*

### **Pendahuluan**

Perkembangan teknologi informasi yang semakin pesat di era globalisasi saat ini semakin memberikan kemudahan untuk berkomunikasi, contohnya adalah internet. Menurut Kementerian Komunikasi dan Informatika[1] pengguna internet di Indonesia saat ini mencapai 63 juta orang. Dari angka tersebut, terdapat 95 persennya menggunakan internet untuk mengakses media sosial. Salah satu media sosial yang diakses masyarakat adalah Twitter. Twitter sebagai salah satu situs microblogging mengalami banyak perkembangan, terhitung pada 21 Maret 2016, Twitter genap memasuki usianya yang ke 10.

Media sosial ini secara global memiliki 332 juta pengguna bulanan dengan 500 juta kicauan dikirim setiap hari dan 200 miliar kicauan dalam setahun [2]. Hal ini tentu saja menyebabkan semakin banyaknya pengguna

diseluruh dunia yang berbagi dan membuat posting tentang produk dan layanan yang mereka gunakan atau mengekspresikan pandangan mereka tentang isu agama dan politik.

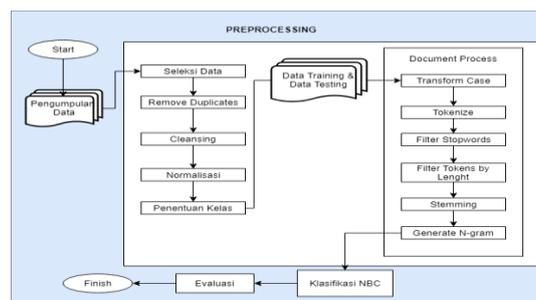
Twitter juga tidak jarang dimanfaatkan oleh beberapa orang untuk memudahkan pekerjaannya, sebagai contoh seorang gubernur yang menggunakan twitter sebagai media untuk mendengarkan keluhan keluhan – keluhan dan masukan dari warganya. Dan di Indonesia sendiri hal tersebut sudah dilakukan oleh beberapa kepala daerah yang aktif sekali bermain mikro blogging Twitter ini. Ridwan Kamil yang sampai saat ini telah memiliki 3,6 juta pengikut atau followers di akun Twitter miliknya yaitu @ridwankamil memanfaatkan Twitter dengan baik untuk menjaga kedekatan dengan warganya. Namun pemerintahannya sendiri tidak luput dari komentar publik dan menuai berbagai macam komentar[3]. Dengan banyaknya pengguna Twitter yang menyampaikan opini – opini tersebut dapat dimanfaatkan untuk mencari sebuah informasi. Salah satu metode yang dapat digunakan untuk menganalisis opini – opini Twitter adalah analisa sentimen[4].

Penelitian ini dilakukan karna ridwan kamil sendiri mendapatkan banyak sekali dukungan dari masyarakat untuk maju menjadi calon Presiden RI tahun 2024. Selain dukungan Ridwan Kamil sendiri mendapat berbagai macam kritik atau ujaran kebencian. Pandangan tersebut belum diklasifikasikan sehingga masyarakat belum tau berapa banyak tingkat sentimen positif dan negatif terhadap Ridwan Kamil. Hal tersebut tentunya menimbulkan berbagai macam pandangan masyarakat terhadap kepemimpinan Ridwan Kamil sendiri sebagai Gubernur Jawa Barat.

## Metode Penelitian

### Kerangka Berfikir

Langkah-langkah penelitian sesuai dengan alur kerangka berfikir berikut ini:



Gambar 1 Kerangka Berfikir

### Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan proses crawling data twitter menggunakan software Orange 3, data yang dikumpulkan merupakan tweet yang berbahasa Indonesia yang diambil dari akun twitter @ridwankamil yang berisi komentar para pengguna twitter.

title	Author	Content
1	@kompascom	BMKG sebelumnya mencatat sebanyak 72 kali gempa bermagnitudo di bawah 5 yang terja...
2	@inilahkorancor	Ridwan Kamil Dukung Pemanfaatan Teknologi untuk Bisnis di TTG IX/2019 https://t.co/nfZ...
3	@imulii	Cape anjir tadi cleaning area, gegara besok mau ada Ridwan Kamil ke tempat kerja(
4	@Rcka	@dimasdn Oh tidak, saya cinta ridwan kamil ❤️
5	@yaudehsi	Ridwan Kamil kalo jadi anak pertama jadinya
6	@Sky_marryland	@sukiono @CH_chotimah UAS itu sama aja dgn ustad Bikini yg dulu ribut sama Ridwan Ka...
7	@coxarchipela	Kemaren okum tidak berpendidikan banyak bgt yg protes masalah Illuminati masjid ranc...
8	@tempofoto	Gaya Ridwan Kamil dan Istri Bersedapa Saat Bekerja https://t.co/WyAVTgJUDU
9	@sbobet78	Hehree gubernur miskin penghargaan @ridwankamil
10	@davidmillit1	@fahiraids @kumpanan marah si ridwan kamil ntar..lahan duit untuk jabar jadi dikit...ekon...
11	@rss_tester	Ridwan Kamil Dukung Pemanfaatan Teknologi untuk Bisnis di TTG IX/2019 https://t.co/gL8T...
12	@kumpanan	@fahiraids @aniesbaswedan Di sisi lain, menurut Gubernur Jabar @ridwankamil, untuk m...
13	@Pemkot_Suka	Penerapan teknologi merupakan salah satu pendonor pembangunan desa di Jawa Barat, ...
14	@antmegapolit	https://t.co/p9SU4LkFLE
15	@antmegapolit	https://t.co/p9SU4LkFLE
16	@Arcani17	Sigana mah kitu...Trus kmaha kang ridwan kamil??

Gambar 2 Proses Pengumpulan Data

### Seleksi Data

Seleksi Data bertujuan untuk menyeleksi data tweet yang akan digunakan nantinya, dikarenakan data tweet yang didapatkan dari proses *crawling* menggunakan Orange 3 belum sepenuhnya bersih dan memiliki data yang tidak diperlukan [5].

### Remove Duplicates

Pada proses ini bertujuan untuk memfilter data *tweet* yang sama, hal ini disebabkan karena twitter memiliki fitur *retweet* yang menyebabkan banyaknya *tweet* yang sama *directweet* oleh pengguna lain dan menimbulkan teks berulang dengan topik yang sama[6][7].

### Cleansing

*Cleansing* bertujuan untuk membersihkan data dari *noise*. Seperti karakter HTML, ikon emosi, hashtag, username, url dan email[8].

### Normalisasi

Mengganti kata yang tidak baku Kata tidak baku merupakan kata yang tidak terdapat pada Kamus Besar Bahasa Indonesia (KBBI)[9].

### Penentuan Kelas Atribut

Data yang sudah melalui proses pada tahapan *preprocessing* kemudian akan dilakukan penentuan kelas pada atributnya[10], penentuan ini diberikan sesuai subjektivitas peneliti, pembagian atribut pada penelitian ini sendiri hanya akan dibagi menjadi positif dan negatif.

### Document Process

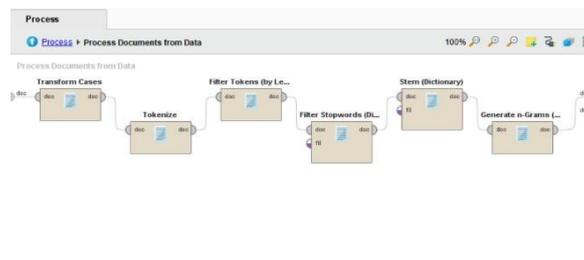
1. *Transform Case*  
Tahapan *transform cases* sendiri dalam penelitian ini bertujuan untuk menyeragamkan bentuk huruf menjadi *lower case* karena pada komentar *Twitter* sendiri memiliki bentuk huruf yang beragam
2. *Tokenize*  
Tokenize merupakan sebuah proses untuk memisahkan atau pemisahan dokumen menjadi per kata atau beberapa kata dalam suatu token/unit kecil.
3. *Filter Token by Length*  
Filter Tokens by Length merupakan proses penghapusan kata atau token berdasarkan panjang karakter yang ditentukan dan pada penelitian ini minimum panjang karakter yang akan digunakan adalah empat karakter pada setiap kata.
4. *Filter Stopwords*  
Tahapan *filter stopword (dictionary)* merupakan tahapan proses yang bertujuan untuk menghapus atau memfilter kata yang tidak memiliki arti dan tidak berhubungan dengan ucapan atau kata *sentiment* [11]

### Stemming

Tahap Stemming adalah tahap menghilangkan imbuhan yang masih melekat sehingga dapat diperoleh kata dasar, dalam penelitian ini proses stemming akan dilakukan dengan menggunakan data stem nolimit.

### Generate N-Grams

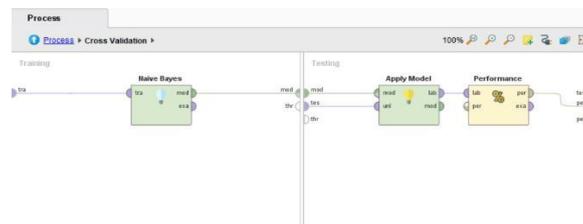
Proses ini merupakan proses pemotongan string atau dokumen (n-gram) berdasarkan limit angka tertentu[12]. Dan dalam penelitian ini sendiri n-gram yang akan digunakan adalah bigram. Tujuan penggunaan N-gram sendiri karena dalam Bahasa Indonesia banyak frase yang tidak hanya terdiri dari satu kata.



Gambar 3 Document Process

### Klasifikasi

Dalam penelitian ini algoritma yang digunakan adalah *Naïve Bayes Classifier* dan berikut merupakan tampilan penerapan algoritma tersebut didalam sebuah sub proses pada operator *cross validation* dengan menambahkan [13],[14] *Naïve Bayes* pada tab *training* dan pada tab *testing* akan ditambahkan *apply model* dan *performance* yang nantinya akan digunakan sebagai penghitung performa klasifikasi.



Gambar 4 Klasifikasi NBC

### Evaluasi

Dalam penelitian ini proses evaluasi akan menggunakan *Confusion Matrix*, Evaluasi bertujuan untuk mengetahui kinerja dari algoritma *Naïve Bayes* yang nantinya akan dijadikan sebagai acuan dari perbandingan akurasi pengujian [15],[16].

Tabel 1 Confusion Matrix

Predicted	Actual Class	
	Class 1	Class 2
	Class 1	True Positive
Class 2	Flase Positive	True Negative

### Hasil Dan Pembahasan

#### Dataset

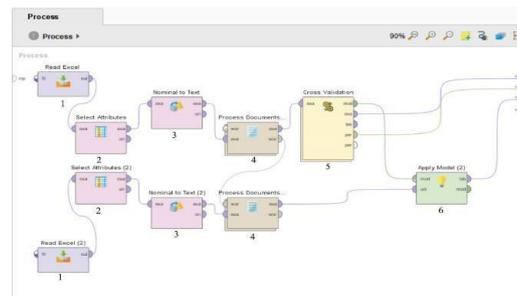
Dataset yang digunakan dalam penelitian ini sebanyak ada 2500 tweet yang kemudian hanya dipilih sebanyak 1000 tweet yang masing-masing diambil dari pengguna yang memention ke akun @ridwankamil dan data diambil dari tanggal yang berbeda-beda dikarenakan twitter sendiri membatasi pengambilan data atau crawling selama 7 hari dari waktu pengambilan data. Untuk lebih jelasnya dapat dilihat pada tabel aberikut:

Tabel 2 Dataset

Dataset	Data Training	Data Testing	Total
@ridwankamil	800	200	1000

#### Proses Pengujian

Proses Pengujian dilakukan dengan mengikuti alur yang telah ditentukan yaitu dengan metode klasifikasi Naïve Bayes dengan Cross Validation, proses ini terdiri dari 10 kali proses pengujian yang mana nantinya akan dijadikan tolak ukur dari penelitian ini. Dalam proses ini peneliti menggunakan RapidMiner untuk melakukan pengujian dengan merangkai beberapa proses yaitu read excel, Select Attribut, Nominal to text, Process Document, Cross Validation dan Apply Model. Proses tersebut dapat dilihat pada gambar dibawah ini:



Gambar 5 Proses Pengujian

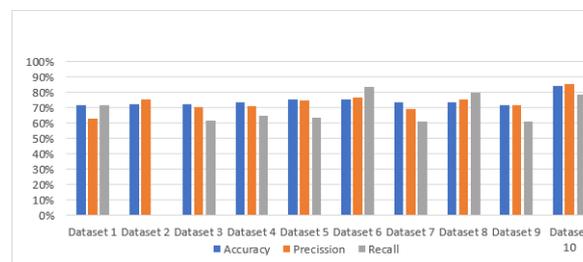
### Analisa Akurasi

Dari hasil pengujian yang telah dilakukan menggunakan algoritma *Naïve Bayes* dan dilakukannya pengujian menggunakan teknik *cross validation* penulis mendapatkan hasil pengujian seperti pada tabel berikut:

Tabel 3 Analisa Akurasi

Dataset	Accuracy	Precision	Recall
Dataset 1	71.62%	62.58%	71.69%
Dataset 2	72.35%	75.14%	78.28%
Dataset 3	72.35%	70.53%	61.78%
Dataset 4	72.35%	70.93%	65.00%
Dataset 5	73.59%	74.67%	63.36%
Dataset 6	75.50%	76.93%	83.32%
Dataset 7	73.22%	68.90%	61.13%
Dataset 8	73.47%	75.07%	79.83%
Dataset 9	71.50%	71.31%	61.14%
Dataset 10	84.38%	85.62%	78.75%

Dari tabel diatas maka dapat disimpulkan bahwa nilai Akurasi , *Precision* dan *recall* tertinggi terdapat pada pengujian ke 10 atau dataset 10 yang mana mendapatkan hasil paling tinggi dengan nilai *Accuracy* 84.38% *Precision* 85.62% dan *Recall* sebanyak 78.75%. Hal ini menunjukkan bahwa penggunaan teknik *cross validation* cukup berpengaruh pada hasil pengujian yang didapatkan.



Gambar 6 Diagram Akurasi

### Preference Value

Preference Value bertujuan untuk mengetahui respon positif masyarakat dari data yang telah diambil dan di proses sebelumnya, jumlah data positif pada data training dan testing berjumlah 583 dan jumlah data negatif pada data training dan testing berjumlah 417, total dari data tersebut berjumlah 1000 data, dan berikut adalah perhitungan preference value dari data tersebut dikalikan dengan nilai akurasi terbaik:

$$\text{Preference Value} = \frac{583}{583+417} * 84.38\% = 0,4919354$$

Semakin tinggi nilai preference value yang dihasilkan maka dapat diartikan semakin tinggi pula respon positif yang didapatkan[17],[18],[19]. Dan dari perhitungan diatas maka dapat disimpulkan bahwa respon positif masyarakat terhadap kepemimpinan ridwan kamil berdasarkan data yang diambil dari Twitter adalah 49%.

## Kesimpulan

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik kesimpulan bahwa penggunaan algoritma Naïve Bayes disertai dengan penggunaan teknik Cross Validation menghasilkan hasil yang cukup baik. Dengan nilai akurasi sebesar 84.38%. dan tingkat respon positif masyarakat terhadap kepemimpinan Ridwan Kamil sebesar 49%. Sedangkan nilai respon positif tersebut dapat berubah-ubah, dikarenakan respon masyarakat dan data yang diperoleh dapat berubah sewaktu-waktu seperti data yang sekarang digunakan peneliti miliki.

## Daftar Rujukan

- [1] Kementerian Komunikasi dan Informatika Republik Indonesia (Kominfo). (2013). Pengguna Internet di Indonesia 63 Juta Orang. Jakarta: *Kementerian Komunikasi dan Informatika Republik Indonesia*.
- [2] Achsa, H. P. (2018). Penggunaan Internet Sebagai Public Sphere dalam Demokrasi Deliberatif (Analisis Penggunaan Hashtag Terkait Isu Politik Menjelang Pilpres 2019) (*Doctoral dissertation, University Of Muhammadiyah Malang*).
- [3] R. S. Putra, (2017). "Analisis sentimen twitter dengan klasifikasi naïve bayes menggunakan seleksi fitur mutual information dan inverse document frequency riky sutriadi putra,"
- [4] W. C. Indhiarta, (2017). "Penggunaan N-Gram Pada Analisa Sentimen,"
- [5] Prabowo, T. T. (2020). Memperebutkan ruang publik virtual: literasi, hoax, dan perdamaian. *Zabir Publishing*.
- [6] Damaris, Y., Sihabudin, A., & Praceka, P. A. (2016). Trending Topic Twitter dalam Menentukan Agenda Pemberitaan di Media Konvensional (Studi Kasus Terhadap Pemberitaan di Kompas TV Periode Oktober-Desember 2015) (*Doctoral dissertation, Universitas Sultan Ageng Tirtayasa*).
- [7] Suherman, S., & Siregar, R. M. S. (2020). Klasifikasi Analisis Sentimen Terhadap Calon Presiden 2019 Pada Media Sosial Twitter Menggunakan Metode Algoritma Naïve Bayes. *Jurnal SIGMA*, 10(3), 15-25.
- [8] Firmansyah, A., & Eriswandi, E. (2020). Analisis Sentimen Terhadap Operator Seluler Telkomsel Menggunakan Algoritma Naïve Bayes. *Jurnal SIGMA*, 10(4), 108-116.
- [9] Fitriani, A. M., Mahardika, E. F., Maulana, M. Y., & Ulya, C. (2021). Analisis Kesalahan Berbahasa Tataran Ejaan pada Majalah "Mulia" Edisi April 2020. *Jurnal Genre*, 3(1), 1-9.
- [10] Ripai, M., Hayati, U., Widyawati, W., & Susana, H. (2022). Klasifikasi Surat Pemberitahuan Pajak Daerah Menggunakan Metode Regresi Logistik Biner Untuk Mengetahui Patuh Dan Tidak Patuh Dalam Pembayaran Pajak Daerah. *KOPEKTIP: Jurnal Ilmiah Manajemen Informatika dan Komputer*, 6(1), 27-33.
- [11] Ginting, H. S., Lhaksmana, K. M., & Murdiansyah, D. T. (2018). Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa Barat 2018 di Twitter Menggunakan Naive Bayes. *eProceedings of Engineering*, 5(1).
- [12] Hasanudin, (2018), "Analisa sentimen terhadap tokoh publik pada twitter menggunakan metode klasifikasi naive bayes classifier, term frequency, tf-idf dan n-gram," p. 49.
- [13] G. A. Buntoro, (2019), "Analisis Sentimen Calon Gubernur Jawa Timur 2018 di Twitter," *Sci. Comput. Sci. Informatics J.*, vol. 1, no. 2, p. 37.
- [14] Romli, I., & Putra, B. M. (2020). Evaluasi Penilaian Kinerja Dalam Klasifikasi Data Mining Dengan Metode Naive Bayes. *Jurnal Teknik Industri*, 1(2), 36-45.
- [15] Kusuma, A., & Nugroho, A. (2021). Analisa Sentimen pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes. *Jurnal Ilmiah Teknologi Informasi Asia*, 15(2), 137-146.
- [16] Romli, I., Pardamean, T., Butsianto, S., Wiyatno, T. N., & bin Mohamad, E. (2021, March). Naive bayes algorithm implementation based on particle swarm optimization in analyzing the defect product. In *Journal of Physics: Conference Series* (Vol. 1845, No. 1, p. 012020). IOP Publishing.
- [17] Wiyanto, W., & Setyaningsih, Z. (2021). Sentiment Analysis Pemutusan Hubungan Kerja Akibat Pandemi Covid-19 Menggunakan Algoritma Naïve Bayes Dan PSO. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 10(3), 426-431.
- [18] Ahmad Turmudi Zy, (2017), "Comparison Algorithm Classification Naive Bayes, Decision Tree, And Neural Network For Analysis Sentiment" *Progr. Stud. Tek. Inform. Sekol. Tinggi Teknol. Pelita Bangsa*, vol. 12, no. April, p. 200.
- [19] H. S. Ginting, K. M. Lhaksmana, and D. T. Murdiansyah, (2018), "Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa Barat 2018 di Twitter Menggunakan Naive Bayes," *ISSN 2355-9365 e-Proceeding Eng. Vol.5, No.1 Maret 2018, Univ. Telkom*, vol. 5, no. 1, pp. 1793–1802.