

# **Analisis Sentimen Pemilihan Gubernur Jawa Barat Tahun 2018 Dengan Aplikasi Twitter Menggunakan Metode *Naïve Bayesian Classification***

**Yusran Tarihoran dan Kevin Jeremy Manurip  
Fakultas Teknologi Informasi, Universitas Advent Indonesia**

## **Abstrak**

Pemilihan Gubernur Jawa Barat 2018 ramai diperbincangkan di dunia nyata maupun dunia maya, khususnya di media sosial Twitter. Semua orang bebas berpendapat atau beropini tentang calon Gubernur Jawa Barat 2018 sehingga memunculkan banyak opini, tidak hanya opini yang positif atau netral, adapula 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 salah satu tokoh publik yang mencalonkan diri gubernur Jawa Barat. Metode penelitian yang digunakan dalam riset kali ini adalah algoritma klasifikasi *Naïve Bayesian Classifier*. Data yang digunakan adalah tweet berbahasa Indonesia dengan kata kunci Ridwan Kamil (#RidwanKamil) sebanyak 1031 data tweet selamat setiap hari dimulai dari 15 Januari 2018 sampai 15 April 2018. Hasil dari klasifikasi menggunakan algoritma *Naïve Bayesian Classifier* didapat 690 jumlah tweet atau 67% dari jumlah keseluruhan data tweet yang mendukung bapak Ridwan Kamil atau bersifat positif khususnya terhadap program kerja yang akan dilakukan dan ini memberikan statistik probabilitas sebesar 73,13% tingkat akurasi *Correctly Classified Instances*.

**Kata – kata kunci:** analisis sentimen, calon gubernur jawa barat, ridwan kamil, *naïve bayesian classification*.

## **Sentiment Analysis of 2018 West Java Governor Election in Twitter Application Using the *Naïve Bayesian Classification* Method**

### **Abstract**

*Elections on West Java Governor 2018 are busy discussed in the real world and cyberspace, especially in social media Twitter. Everyone is free to argue about the candidate of West Java Governor 2018 that raises many opinions, not only positive or neutral opinion, but also negative opinion. Nowadays, social media especially Twitter is one of the place to promote or to campaign effectively and efficiently to increase supporters interest. In this case researchers will conduct research on one of the public figures who run for governor elections West Java. The research method used in this research is the *Naïve Bayesian Classifier Classification* algorithm. The data used is an Indonesian tweet with the keyword Ridwan Kamil (#RidwanKamil) as much as 1031 data tweet every day starting from January 15, 2018 to April 15, 2018. Results from the classification using the *Naïve Bayesian Classifier* algorithm obtained 690 number of tweets or 67% all data tweets that support Mr. Ridwan Kamil or are positive especially on the work program that will be done and this provides a probability statistics of 73.13% accuracy level *Correctly Classified Instances*.*

**Keywords:** *Sentiment Analysis, Candidate of West Java Governor, Ridwan Kamil, Naïve Bayesian Classification*

### **Pendahuluan**

Dewasa ini semakin banyak orang menyatakan pendapat atau opini mereka terhadap tokoh-tokoh masyarakat seperti pemimpin wilayah, pejabat, orang berpengaruh, dan sebagainya. Hal ini dimanfaatkan oleh para peneliti untuk mengetahui pendapat orang mengenai siapa orang-orang tersebut. Sebaliknya bagi produsen dapat digunakan sebagai media pelacak atau tolok ukur tentang seberapa besar parameter dari orang-orang tersebut di mata produsen. 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.

Menurut (Putranti & Winarko, 2014) analisis sentimen atau sering disebut dengan *opinion mining* adalah penentuan dibalik serangkaian kata-kata, yang digunakan untuk memperoleh pemahaman tentang sikap, pendapat dan emosi yang diekspresikan secara *online*. Analisis sentimen sangat berguna dalam pemantauan media sosial karena memungkinkan kita untuk memperoleh gambaran tentang opini publik yang luas dari suatu topik-topik tertentu. Peran media sosial seperti Twitter dan Facebook, saat ini menjadi sebuah media atau ruang yang memberikan kritik terhadap kegiatan politik. Dalam hal ini, peneliti menganalisis opini, sentimen, evaluasi, emosi, serta sikap terhadap sebuah entitas seperti peristiwa topik dan atribut pada subjek. Dalam penelitian kali ini, penulis menggunakan metode *Naïve Bayesian Classification* untuk menganalisis permasalahan yang terjadi.

## Landasan Teori

Ketika suatu organisasi/perusahaan/perorangan ingin memperoleh opini publik mengenai produk, citra, dan layanannya, maka mereka tidak perlu melakukan survei konvensional dan fokus grup yang mahal biayanya. Web melalui situs review online, blog pribadi, situs jejaring sosial lainnya menyediakan sumber-sumber opini yang besar jumlahnya bagi kebutuhan individu maupun organisasi. Melalui web, orang dapat mengekspresikan apa saja, termasuk pendapatnya akan suatu hal tanpa adanya keterpaksaan (Buntoro, Adji, & Purnamasari, 2014).

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.

Jika dipandang dari kacamata demokrasi, tujuan pemilu hendkalah kembali berpegang pada prinsip kebijaksanaan yang demokratis yaitu menjadkan kepentingan semua golongan masyarakat. Untuk itu tujuan pemilu harus dinyatakan dalam fungsi-fungsi utama pemilu yaitu:

1. Membentuk pemerintahan perwakilan lewat partai politik pemenang pemilu.
2. Menentukan wakil rakyat di lembaga perwakilan rakyat.
3. Pergantian atau pengukuran elit penguasa.
4. Pendidikan politik bagi rakyat melalui partisipasi masyarakat di dalam pemilu.

Menurut dari (Lidya, Sitompul, & Efendi, 2015), penelitian yang mereka lakukan adalah mengklasifikasikan teks Bahasa Indonesia yang terdapat di situs web berupa artikel berita, kemudian menggunakan metode *K-Nearest Neighbor* akan mengklasifikasi langsung pada data pembelajaran agar dapat menentukan model yang dibentuk oleh metode *Support Vector Machine* untuk menentukan kategori dari data baru yang ingin ditentukan kategori tekstual, yaitu sentimen positif, sentimen negatif, dan sentimen netral.

Dalam hal ini penelitian dilakukan dengan menggunakan metode *Support Vector Machine*. Hubungan antara metode ini dengan text mining memiliki keterkaitan yang sangat erat terlebih dalam melakukan pengklasifikasi informasi yang masuk. menurut menurut (Wikipedia, 2017), text mining merupakan proses ekstraksi pola yang berupa informasi dan pengetahuan dari sejumlah besar data. Jenis data masukan/input data merupakan data tak terstruktur.

Menurut Taufik Hidayatulloh (Hidayatulloh, 2015) dalam penelitiannya mengenai Kajian Komparasi Penerapan Algoritma *Support Vector Machine* (SVM) dan *Multilayer Perception* (MLP) dalam Prediksi Indeks Harga Saham Sektor Perbankan: Studi Kasus Saham LQ34 IDX Bank BCA mengungkapkan pendapatnya mengenai penggunaan SVM sebagai metode klasifikasi yang memadai dalam text mining khususnya untuk menganalisa Indeks Harga Saham (IHS). Hasil dari Analisa tersebut menyimpulkan bahwa penerapan algoritma *Support Vector Machines* (SVM) 10 lag menghasilkan keputusan yang lebih akurat dibandingkan dengan algoritma Multilayer Perception (MLP) baik dari lag 6 maupun lag 10 dengan nilai RMSE (Root Mean Squared Error) paling kecil 317,89, ME = 639,04, MAE = 262,72, dan MAPE = 1,81 dalam

meramalkan/memprediksikan indeks harag saham perbankan (BBCA.JK). Dengan ini, makan algoritma SVM mendapat kategori terbaik dalam meramalkan/memprediksian indeks harga saham. Sedangkan algoritma MLP hasil dari setiap model sangat bergantung pada pemilihan jumlah variable prediksi, topologi jaringan, fungsi aktivasi, dan metode pelatihan jaringan.

## Metodologi Penelitian

### Crawling Data

Dari pembahasan bab sebelumnya telah dijabarkan tentang teori-teori mengenai algoritma-algoritma yang melakukan kaslifikasi terhadap 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 Ridwan Kamil 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.

### Mengumpulkan Data *tweet*

Data tweet diambil dengan metode Crawling dari media sosial Twitter. Data yang akan diambil hanya tweet dalam bahasa Indonesia dimulai dari 4 bulan sebelum pemilu pilgub 2018 secara real-time (pengambilan setiap hari) yaitu 1031 tweet dengan kata kunci #RidwanKamil. Data diambil secara acak baik dari user biasa ataupun media online di Twitter.

**Tabel 1** Contoh data tweet

No.	Tweet	Sentimen
1.	Ridwan Kamil katanye gagal mengatasi masalah utama Bandung, hm.. 😞	Negatif
2.	Ridwan Kamil sering dipuji sukses tapi warga bandung masih banyak buang air besar sembarangan	Negatif
3.	Jabar juara ridwan kamil.	Positif
4.	Jabar 1 Ridwan Kamil kerja nyata. hwaw	Positif
5.	Yuk politik mungkin saja dapat sepeda dari toko.	Netral
6.	Kampanye di Bekasi, Ridwan Kamil menantang wartawan untuk adu panco.	Netral

### Tahap *Preprocessing Data*

Pada tahap pre-processing data, dilakukan 4 langkah sebagai berikut.

1. Seleksi Komentar  
Tahap ini, dilakukan seleksi komentar yang mengandung kata atau hashtag(#) Ridwan Kamil, karena Twitter terdapat fungsi retweet, yaitu memberikan komentar terhadap tweet komentar seseorang, karena komentar seseorang, karena komentar tweet akan mengganggu dalam proses Analisis Sentimen tweet, jadi dalam pre-processing ini komentar tweet dihapus.
2. Cleansing

Kalimat yang didapat biasanya masih terdapat noise, yaitu kesalahan acak atau varian dalam variabel terukur, untuk itu, kita harus menghilangkan noise tersebut. Kata yang dihilangkan adalah karakter HTML, kata kunci, ikon emosi, huruf besar, tanda baca, hashtag (#), username (@username), url (https://website-name.com), dan email (nama@website.com).

3. Parsing  
Yaitu proses memecahkan dokumen menjadi sebuah kata dengan melakukan analisis terhadap kumpulan kata dan memisahkan kata tersebut lalu menentukan struktur sintaksi dari tiap kata tersebut.
4. Normalisasi Kata  
Bertujuan untuk menormalkan kalimat sehingga kalimat gaul/alay/singkat menjadi normal, sehingga Bahasa-bahasa tersebut dapat dikenali sebagai Bahasa yang sesuai dengan KBBI. Yang harus dilakukan untuk melakukan normalisasi kalimat adalah:
  - Meregangkan tanda baca (punctuation) dan symbol selain alphabet.
  - Meregangkan tanda baca adalah memberikan jarak terhadap tanda baca dari kata-kata sesudah atau sebelumnya, tujuannya agar tanda baca dan symbol selain alphabet tidak menjadi satu dengan kata-kata pada saat proses tokenizing.
  - Mengubah menjadi huruf kecil untuk semua kalimat dan kata.
  - Normalisasi kata.
5. Menghilangkan huruf berulang  
Ketika sedang senang atau kesal, seseorang bebas menuliskan opini berdasarkan emosinya, biasanya seseorang menuliskan kesenangan. Kata berulang seperti "kereeen" akan di normalisasi menjadi "keren".
6. Menghilangkan emoticon  
Ketika sedang menulis status (tweet) seseorang kadang salah satu kurang tepat dalam penggunaan emoticon, entah disengaja atau tidak banyak yang melakukannya. Jadi, dalam proses ini emoticon dihapus atau diabaikan agar hasil tidak menjadi data bisa.

### Tokenisasi

Setelah melakukan normalisasi kata dan kalimat, selanjutnya kalimat tersebut dipecah ke dalam token-token menggunakan pembatas atau delimiter spasi. Token yang digunakan penelitian ini yaitu: N-gram: token yang berdiri dari tiga kata setiap satu token, contohnya: Daerah Khusus Ibukota.

**Tabel 2** Tokenisasi

Tweet	Tokenizing
Yuk	mari
Utk	untuk
Judul	Judul
Ridwan	ridwan
Kamil	kamil
"Semut	semut
Di Sebrang	di sebrang
Tampak, Gajah	tampak, gajah
Di Pelupuk Mata"	di pelupuk mata

### 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, karena dalam dunia maya komentar yang dilontarkan oleh netizen mengandung unsur positif, netral, dan negatif.

### Penerapan Metode *Naïve Bayesian Classification*

Menurut (Ariadi & Fithriasari, 2015) Metode Naïve Bayes Classification (NBC) merupakan salah satu metode yang dapat mengklasifikasikan teks. Kelebihan NBC adalah dalam setiap dokumen direpresentasikan dengan pasangan atribut "a1, a2, ax ... an" di mana a1 adalah kata pertama, a2 adalah kata kedua, dan seterusnya. Sedangkan V adalah himpunan kategori berita. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (Vmap). Adapun Vmap adalah sebagai berikut

Dimana |doc| merupakan jumlah dokumen (artikel berita) yang memiliki kategori j dalam training. Sedangkan |training| merupakan jumlah dokumen (artikel berita) dalam contoh yang digunakan untuk training. Untuk probabilitas kata a1 untuk setiap kategori P(ai|vj), dihitung pada saat training (Ariadi & Fithriasari, 2015). Dimana ni adalah jumlah kemunculan kata a1 dalam dokumen yang berkategori vj, sedangkan n adalah banyaknya seluruh kata dalam dokumen dengan kategori vj dan |kosakata| adalah banyaknya kata dalam contoh pelatihan (Ariadi & Fithriasari, 2015).

### Hasil Penelitian dan Impelementasi

a	b	c	<-- classified as
690	0	0	a = Positif
162	56	0	b = Netral
115	0	8	c = Negatif

**Gambar 1** Confusion Matrix

*Correctly Classified Instances* merupakan suatu parameter dari tingkat akurasi yang benar (74,9% = 754 data *tweet*) dan tingkat dari kesalahan atau disebut dengan *Incorrectly Classified Instances* (26,8% = 277 data *tweet*) yang bisa kita analisa dengan:

$$\frac{\text{Jumlah status yang diklasifikasi benar dan memang benar}}{\text{Jumlah status (tweet) secara keseluruhan}} \times 100\% \\ \frac{754 \text{ tweet}}{1031 \text{ tweet}} \times 100\% = 0.731 = 73,1\%$$

Sedangkan untuk *Incorrectly Classified Instances* berikut dengan perhitungannya

$$\frac{\text{Jumlah status yang diklasifikasi salah dan memang salah}}{\text{Jumlah status (tweet) secara keseluruhan}} \times 100\% \\ \frac{277}{1031} \times 100\% = 0.2679 = 26.8\%$$

Penjelasan *Confusion Matrix*:

1. Baris pertama "690 - 0 - 0" atau 690 *tweet* yang mendukung dan data sebenarnya mendukung.
2. Baris kedua "126 - 56 - 0" atau 182 *tweet* yang menunjukkan bahwa 126 *tweet* mendukung tetapi masih bersifat tidak mutlak, 56 *tweet* bersifat netral atau tidak memihak kepada siapa-siapa.
3. Baris ketiga "115 - 0 - 8" atau 123 *tweet* yang menunjukkan bahwa 115 *tweet* mendukung tetapi sebenarnya menolak dan bersifat tidak mutlak, dan 8 *tweet* bersifat negatif atau menolak secara mutlak.

### Kesimpulan dan Saran

#### Kesimpulan

Berdasarkan analisis dan pengujian yang telah dilakukan pada bab 4, maka kesimpulan yang pengujian dapat diambil adalah sebagai berikut:

1. Algoritma klasifikasi yang digunakan, yaitu Naïve Bayesian Classifier dapat meningkatkan efisiensi sebesar 73,13% parameter akurasi sehingga memaksimalkan hasil dari penentuan data yang diproses.
2. Dari hasil klasifikasi Correctly Classified Instances berjumlah 73.13% terdapat 690 jumlah tweet atau 67% Mendukung Ridwan Kamil khususnya terhadap program kerja yang telah beliau buat untuk visi kedepannya.
3. Dengan menggunakan 1031 tweet data train sebagai bahan analisa maka akan semakin akurat untuk mengujinya, karena semakin banyak data training yang digunakan maka akan semakin banyak kamus kata yang tersimpan. Sehingga semakin banyak pula kata-kata yang bisa diakses untuk menjadi parameter sentimen di media sosial.

### Saran

Penulis menyarankan pengembangan penelitian lebih lanjut dengan sistem pengklasifikasian data tweet sebagai berikut:

1. Penulis berharap penelitian selanjutnya semua klasifikasi yang telah dilakukan mendapatkan fitur yang lebih baik lagi untuk ke depannya, seperti penggunaan Part of Speech tagging untuk mengetahui posisi sebuah kata dalam kalimat.
2. Bahasa yang digunakan tidak hanya Bahasa Indonesia saja, tetapi bisa menggunakan Bahasa daerah atau Bahasa asing, karena bertujuan untuk menjangkau data tweet lebih banyak lagi.
3. Penelitian ini dapat dimanfaatkan oleh orang-orang sebagai sebuah parameter untuk menentukan pilihan atau sentimen mereka terhadap tokoh publik yang menjadi objek penelitian.

### Referensi

1. Ariadi, D., & Fithriasari, K. (2015). Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer. JURNAL SAINS DAN SENI ITS Vol. 4, No.2, 4(2), 248–253.
2. Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. 32 Integer Journal Maret, 1(2016), 32–41. Retrieved from [https://www.researchgate.net/profile/Ghulam\\_Buntoro/publication/316617194\\_Analisis\\_Sentimen\\_Calon\\_Gubernur\\_DKI\\_Jakarta\\_2017\\_Di\\_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf](https://www.researchgate.net/profile/Ghulam_Buntoro/publication/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf)
3. Buntoro, G. A., Adji, T. B., & Purnamasari, A. E. (2014). Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation. Citee 2014, ISSN: 2085(OCTOBER 2014), 7–8.
4. Feldman, R., & Sanger, J. (2007). The text mining handbook: advanced approaches in analyzing unstructured data. *Imagine*, 34, 410. <https://doi.org/10.1179/1465312512Z.00000000017>
5. H. Hasbullah, H Nasution, M. Saleh, L. P. S. islahiyah B. (2015). Laporan riset partisipasi pemilih pada pemilihan umum.
6. Hidayatullah, A. F., & Sn, A. (2014). Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter. Seminar Nasional Informatika 2014, 2014(August 2013), 0–8.
7. Hidayatulloh, T. (2015). Kajian Komparasi Penerapan Algoritma Support Vector Machine ( Svm ) Dan Multilayer Perceptron ( Mlp ) Dalam Prediksi Indeks Saham Sektor Perbankan : Studi Kasus Saham Lq45 Idx. *Bina Sarana Informatika*, (July).
8. Khodra, M. L., Ayu, P., Insanudin, A., & Megally, M. (2013). Ekstraksi Informasi Transaksi Online pada Twitter. *Cybermatika*, 1(July), 1–4. Retrieved from <http://elib.uni-stuttgart.de/opus/volltexte/2012/8012/>
9. Kompas. (2018). Debat Publik Pertama: Pilgub Jawa Barat. Retrieved from <https://www.youtube.com/watch?v=lbYD8FOKXCA>
10. KPU. (2014). Daftar penelitian komisi pemilihan umum republik indonesia.
11. Lidya, S. K., Sitompul, O. S., & Efendi, S. (2015). Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine ( Svm ). Seminar Nasional Teknologi Dan Komunikasi 2015, 2015(Sentika), 1–8. <https://doi.org/10.1016/j.eswa.2013.08.047>
12. Ling, J., Kencana, I. putu E. N., & Oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, 3(3), 92–99.

13. Mujilawati, S. (2016). Pre-Processing Text Mining Pada Data Twitter. Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016(Sentika), 2089–9815.
14. Pikiran Rakyat. (2018). Tiga Masalah Utama Fokus Ridwan Kamil dan Uu Ruzhanul Ulum. Retrieved from <http://www.pikiran-rakyat.com/bandung-raya/2018/03/13/tiga-masalah-utama-fokus-ridwan-kamil-dan-uu-ruz-hanul-ulum-421216>
15. Putranti, N. D., & Winarko, E. (2014). Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. *Ijcss*, 8(1), 91–100. <https://doi.org/10.22146/ijccs.3499>
16. Raharjo, B. (2018). Sentimen Analisis yang baik dan benar.
17. Saputra, N., Adji, T. B., & Permanasari, A. E. (2015). Analisis Sentimen Data Presiden Jokowi dengan Preprocessing Normalisasi dan Stemming menggunakan Metode Naive Bayes dan SVM, 5(November).
18. 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.
19. Wikipedia. (2017). Penambangan Teks. Retrieved from [https://id.wikipedia.org/wiki/Penambangan\\_teks](https://id.wikipedia.org/wiki/Penambangan_teks)