

ANALISIS SENTIMEN TERHADAP PEMERINTAHAN JOKO WIDODO PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVES BAYES CLASSIFIER

Yonathan Sari Mahardhika¹, Eri Zuliarso²

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Stikubank
e-mail: ¹yonathan.sari@gmail.com, ²eri299@edu.unisbank.ac.id

ABSTRAK

Twitter merupakan media sosial yang sedang populer saat ini, disini publik bebas berkomentar dan menulis apapun. Tidak jarang publik berkomentar dengan kata – kata kasar bahkan ujaran kebencian. Pemerintahan Joko Widodo menuai banyak komentar, ada yang memuji, mengkritik dan menghina. Untuk dapat menggali informasi dan melakukan klasifikasi sebuah teks diperlukan analisis sentimen. Dalam penelitian ini analisis sentimen merupakan proses klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen negatif dan positif. Data opini diperoleh dari jejaring sosial Twitter berupa tweet berdasarkan query dalam Bahasa Indonesia. Data yang digunakan berjumlah 400 tweet terdiri dari 300 data latih dan 100 data uji. Data latih merupakan data yang telah diketahui sentimennya, 300 data latih terdiri dari 150 data kelas sentimen negatif dan 150 data kelas sentimen positif. Penelitian ini bertujuan untuk menentukan tweet merupakan tweet positif atau negatif yang disampaikan di Twitter dalam Bahasa Indonesia. Pengklasifikasian data tweet menggunakan algoritma naïve bayes classifier. Hasil klasifikasi pada data uji menunjukkan, algoritma Naïve Bayes Classifier memberikan nilai akurasi sebesar 97%. Untuk nilai akurasi tiap sentimennya yaitu 96% untuk sentimen negatif dan 98% untuk sentimen positif.

Kata kunci : Analisis Sentimen, Klasifikasi, Naïve Bayes, Twitter, Joko Widodo.

1. PENDAHULUAN

Di era digital ini, media sosial merupakan salah satu tempat atau media untuk berekspresi dan berpendapat tentang berbagai macam topik. Sosial media adalah satu set baru komunikasi dan alat kolaborasi yang memungkinkan banyak jenis interaksi yang sebelumnya tidak tersedia untuk orang biasa (Brogan, 2010). Twitter adalah tempat yang tepat untuk berbagi ide, bank gagasan, tempat untuk mengumpulkan informasi, untuk menginspirasi pikiran, atau untuk melihat apa yang teman lakukan (Brogan, 2010). 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 (Maulana, 2016). Banyak pengguna twitter yang melakukan posting ekspresi dan pendapat mereka terhadap sebuah produk, layanan, isu politik atau hal – hal yang sedang viral. Pemerintahan yang sedang berjalan pada saat inipun tidak luput dari komentar publik dalam media sosial twitter. Pemerintahan saat ini yang dipimpin Presiden Joko Widodo menuai berbagai macam komentar, mulai dari pujian, kritik, saran, sindiran bahkan ujaran kebencian. Dengan banyaknya pengguna Twitter yang menyampaikan opini-opini tersebut dapat dimanfaatkan untuk mencari sebuah informasi. Namun dalam pemanfaatannya membutuhkan analisis yang tepat sehingga informasi yang dihasilkan dapat membantu banyak pihak untuk mendukung suatu keputusan atau pilihan. Salah satu metode yang dapat digunakan untuk menganalisis opini – opini tweet adalah analisa sentimen.

Analisa sentimen atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini (Rozi, Pramono & Dahlan, 2013). Dalam penelitian ini analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek mengandung sentimen negatif atau positif menggunakan algoritma *Naïve Bayes Classifier* (NBC).

Permasalahan yang dihadapi saat ini adalah bagaimana algoritma *naive bayes classifier* dapat melakukan klasifikasi sebuah tweet termasuk ke dalam kelas sentimen negatif atau positif secara otomatis dan seberapa akurat algoritma *naive bayes classifier* dalam melakukan klasifikasi ke dalam sentimen negatif dan positif.

2. TINJAUAN PUSTAKA

Hasil penelitian terdahulu yang dilakukan oleh Dyarsa (2015) analisis sentiment pada sosial media twitter menggunakan *Naives Bayes Classifier* terhadap kata kunci “Kurikulum 2013”, menghasilkan sebuah sistem yang dapat mengklasifikasi sentimen secara otomatis ke dalam kategori positif, negatif dan netral dengan hasil pengujian 3000 data latih dan 100 data uji, tingkat akurasi mencapai 91%.

Penelitian yang dilakukan oleh Faradhillah, Kusumawardani dan Hafidz (2016), pengklasifikasian dengan menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM) pada akun twitter resmi pemerintah kota

Surabaya @e100ss dan @SapawargaSby. Hasil penelitian ini penggunaan algoritma *Naive Bayes* dengan metode pra-proses penghapusan kata (*stopword removal*) mendapatkan hasil yang lebih tinggi dibanding yang tidak menggunakan. Untuk algoritma *Support Vector Machine* (SVM) metode pra-proses penghapusan kata (*stopword removal*) tidak terlalu berpengaruh terhadap hasil akurasi. Model klasifikasi terbaik dengan menggunakan *Support Vector Machine* (SVM) yang dibangun dengan menggunakan data latih menghasilkan klasifikasi dengan nilai akurasi, precision, recall, dan f-measure berturut-turut pada kelas netral sebesar 78,66%, 99,85%, 79,67%, dan 88,6%. Nilai pada kelas netral memiliki nilai yang paling baik diantara positif dan negatif dikarenakan memang kelas netral mempunyai banyak anggota.

Penelitian yang dilakukan oleh Lorosae, dkk (2018), analisis sentimen pada media sosial terkhusus twitter dalam pelayanan yang disediakan oleh penyedia jasa ekspedisi barang merupakan proses untuk mengelompokkan respon seseorang terhadap pelayanan pengiriman barang yang sedang atau telah dilakukan. Permasalahan yang terjadi dalam penelitian ini terdapat kesulitan untuk menentukan opini yang bersifat positif, negatif ataupun netral. Penelitian ini bertujuan untuk mengembangkan aplikasi yang mampu melakukan analisis sentimen konsumen, dimana *Naive Bayer Classifier* digunakan sebagai metode klasifikasi. Sistem yang dibuat mampu mengklasifikasikan sentimen pada twitter kedalam sentimen positif, netral atau negatif. Hasil yang didapatkan dari akurasi *naive bayes* pada data uji positif memperoleh ketepatan 84%.

Penelitian yang dilakukan oleh Hanggara, Akhriza dan Husni (2017), sentimen dapat diketahui dari opini konsumen atas produk yang di sampaikan melalui website yang menjual produk tersebut. Beberapa aplikasi berbasis web untuk mengevaluasi sentimen konsumen sudah tersedia di internet. Analisis sentimen biasanya mengklasifikasi opini mnejadi tiga kelas: Senimen positif, negatif dan netral. Namun demikian aplikasi yang dapat melakukan proses pembuatan training dan testing set (dari data set opini konsumen terhadap produk), menganalisis sesentimen konsumen sekaligus mengukur akurasi hasil analisis secara dinamis masih jarang tersedia. Penelitian ini bertujuan untuk mengembangkan aplikasi yang mampu melakukan analisis sentimen konsumen dengan fasilitas seperti yang disebutkan sebelumnya, dimana *Naive Bayer Classifier* digunakan sebagai metode klasifikasi. Hasil analisis pada testing set yang diperoleh dari Goodreads, Twitter, dan Tokopedia menunjukkan bahwa aplikasi yang dikembangkan mampu mengklasifikasikan opini ke dalam tiga kelas positif, negatif dan netral dengan tingkat akurasi rata-rata 70,43% - 88,03%.

3. METODE PENELITIAN

Metodologi penelitian yang digunakan terdiri dari :

3.1 Metode Pengumpulan Data

Beberapa metode untuk memperoleh data atau informasi dalam menyelesaikan permasalahan. Metode yang dilakukan tersebut antara lain:

a. Studi Kepustakaan

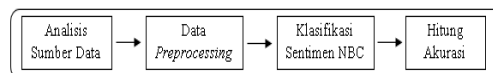
Metode pengumpulan data dengan metode kepubstakaan dilakukan dengan pengumpulan jurnal, literatur, paper, makalah, buku, maupun situs internet sebagai sumber pustaka yang berkaitan dengan materi penulisan khususnya analisis sentimen menggunakan algoritma *naive bayes classifier*.

b. Pengumpulan Data Twitter

Data yang diperoleh merupakan sumber data yang diambil secara langsung dari situs microblogging Twitter dengan keyword pencarian #2019gantipresiden, #2019tetapjokowi dan tweet yang berhubungan dengan pemerintahan pada saat ini yang dipimpin oleh presiden Joko Widodo dengan memanfaatkan search API (Application Programming Interface) yang disediakan oleh Twitter.

3.2 Metode Analisis Data

Metode analisis data merupakan proses yang akan dilakukan untuk menganalisa data tweet sehingga didapat kelas sentimennya. Berikut adalah gambaran penyelesaian masalah dalam penelitian ini dapat dilihat pada gambar 1:



Gambar 1. Alur Sistem Analisis Sentimen

Berikut adalah gambaran penyelesaian analisis sentimen pada gambar 1

a. Analisis Sumber Data

Setelah data didapatkan, langkah selanjutnya ialah melakukan penentuan sentimen secara manual terhadap data tersebut. Data yang diambil berjumlah 400 data tweet terdiri dari 300 data latih dan 100 data tweet. Untuk data latih terdiri dari 150 tweet dengan sentimen negatif dan 150 tweet dengan sentimen positif.

b. Data Preprocessing

Pada tahap ini, tweet yang telah dikumpulkan akan melalui beberapa tahap pemrosesan teks yang terdiri dari case folding, normalisasi fitur, tokenizing dan stopword removal. Tujuan dari pemrosesan teks adalah supaya data yang didapat akan lebih terstruktur agar lebih mudah untuk dilakukan pengolahan data.

c. Klasifikasi Sentimen NBC

Naive bayes classifier adalah penggolongan menggunakan statistik sederhana berdasarkan teorema *bayes* yang mengasumsikan bahwa keberadaan atau ketiadaan dari suatu kelas dengan fitur lainnya (Lorosae, dkk. 2018). Tahap ini adalah proses pengklasifikasian yang akan diproses menggunakan metode *Naive Bayes Classifier* untuk menentukan data mana yang termasuk dalam opini positif dan negatif.

d. Hitung Akurasi

Data yang diuji adalah data yang telah diketahui sentimennya, sehingga setelah dilakuan klasifikasi sentimen maka akan dilakukan perhitungan akurasi atau ketepatan sistem dalam melakukan klasifikasi.

4. HASIL DAN PEMBAHASAN

Tahap awal adalah menyiapkan data *tweet* yang nantinya akan diolah menggunakan algoritma *naive bayes classifier* pada RStudio. Dalam klasifikasi data teks menggunakan algoritma *naive bayes classifier* diperlukan tahap pelatihan dan pengujian. Data *tweet* yang telah didapat akan dibagi menjadi 2 yaitu data latih dan data uji,, pembagian data pada tabel 1.

Tabel 1. Pembagian Data

Jenis Sentimen Tweet	Negatif	Positif
Latih	150	150
Uji	50	50

Setelah menentukan pembagian data latih dan uji, selanjutnya pembacaan data ke dalam RStudio. Gambar 2 adalah partisi data latih, gambar 3 adalah partisi dari data uji.

class	text	username	tgl
1 Neg	Usul yang rasional Dianggap hanya untuk menampung ...	@PembawaCerita	5/17/2018
2 Neg	1. memperburuk hajat hidup orang banyak 2. mengambi...	@DatuakPanduko	5/17/2018
3 Neg	kereeeen mahasiswa medan... Kartu Merah Jokowi... ben...	@DatuakPanduko	5/17/2018
4 Neg	ngomong apa lagi si Jokowi ? 1. Jokowi bilang berantas ...	@DatuakPanduko	5/15/2018
5 Neg	Tak terhitung baryaknya kaum Kafir Harbi pembendI Isla...	@MasPresidenBaru	5/17/2018
6 Neg	Fahri Hamzah: Musuh Kita adalah Pemerintah yang Gag...	@CybermesialD	5/15/2018
7 Neg	#KamiBersamaPOLRI Dolar naik -> bikin berita teroris el...	@DatuakPanduko	5/9/2018
8 Neg	GERGAJI 2018: GERAKAN GANTI NAJIB GERGAJI 2019: GE...	@mpuanon	5/10/2018
9 Neg	wahai umat islam Indonesia, bersatu lah, Berpegang tu...	@mheru212azima	5/17/2018
10 Neg	Nilai rupiah anjlok, artinya @jokowi gagal. Jadi ganti pr...	@kenekmantap	5/9/2018
11 Neg	Woy @jokowi ?? Ads Apa Dengan Pemerintahmu?? Sa...	@SyahAlfath	5/16/2018
12 Neg	Lengserkan Jokowi segera...	@delilapilli	5/18/2018

Gambar 2. Data Latih

class	text	username	tgl
343 Neg	SEMUA ORANG BISA CUMA MENCACI REZIM JOKOWI, PR...	@malik_fadillah	7/11/2018
344 Neg	Duet ASU ini setelah Jokowi terpilih lg di 2019 sebaikny...	@Foya_2	7/9/2018
345 Neg	Fitnah keji.. Klo rezim ini gak di lengserkan, dan dibung...	@PangeranBiru212	7/9/2018
346 Neg	Kita jihad di jalan yg benar.. ALLAHU AKBAR!! #2019Ganti...	@cemjambak	7/8/2018
347 Neg	harus lengserkan Jokowi 2019 nanti .. agar para elit tdk ...	@RI_NDet	6/21/2018
348 Neg	JOKOWI LANGGAR 3 UU DPR HARUS LENGSEKAN JOKO...	@zainal2412	6/19/2018
349 Neg	Mulut lonte, begitu kok mau 2 periode, rakyat semakin ...	@Naay3Dnkri	6/19/2018
350 Neg	Rusak !!! Lengserkan Jokowi ! #2019Retapjokowi ngimpi	@DemonstranTL	6/19/2018
351 Pos	#350MegaWatt #Listrik #SUMETERA #PLTMG #ProgramP...	@irawanandr24	7/27/2018
352 Pos	Inilah alasan utama Kenapa mereka mati2an mau ganti p...	@PalingKafir	7/27/2018
353 Pos	Presiden Jokowi : Pemerintah Sejak Tahun 2015 Telah Dik...	@WidionRizma	7/27/2018
354 Pos	Alasan sederhana mengapa aku pilih #2019Retapjokowi ...	@CaruBiruAtm	7/27/2018

Gambar 3. Data Uji

Kemudian dilakukan pelatihan dan pengujian klasifikasi data. Dari 100 data uji diperoleh hasil klasifikasi prediksi kelas sentimen *tweet* pada gambar 4

```
> predict(classifier, newdata=testNB)
[1] Neg Neg Neg Neg Neg Neg Neg Neg Neg Neg Neg Neg Neg Neg Neg Neg
[19] Neg Neg Neg Neg Neg Pos Neg Neg Neg Neg Neg Neg Neg Neg Neg Neg
[37] Neg Neg Neg Neg Neg Neg Neg Neg Neg Neg Neg Neg Neg Pos Pos Pos Pos
[55] Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos
[73] Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos Pos
[91] Pos Pos Pos Pos Pos Pos Neg Pos Pos Pos
Levels: Neg Pos
```

Gambar 4. Hasil Klasifikasi

Dari pengujian 100 data tweet pada data uji yang terdapat pada gambar 4, algoritma *naive bayes classifier* memberikan prediksi kelas sentimen data dengan rincian 49 dokomune sentimen negatif dan 51 dokumen sentimen positif. Dari hasil prediksi akan dihitung tingkat akurasi algoritma *naive bayes classifier* dalam melakukan klasifikasi *tweet*. Untuk menghitung tingkat akurasi maka digunakan *confusion matrix*. Data hasil prediksi akan dibandingkan dengan aktual kelas. Hasil dari *confusion matrix* pada gambar 5.

Predictions	Actual	
	Neg	Pos
Neg	48	1
Pos	2	49

Gambar 5. Confusion Matrix

Dari gambar 3.4 di dapatkan hasil :

- a. True Positif 49 dokumen
- b. True Negatif 48 dokumen
- c. False Positif 1 dokumen
- d. False Negatif 2 dokumen

Dengan hasil dari confusion matrix maka didapatkan akurasi sebesar

$$\frac{TN+TP}{TN+TP+FN+FP} = \frac{48+49}{48+49+2+1} = \frac{97}{100} = 0,97$$

Hasil akurasi algoritma naive bayes classifier sebesar 0,97 atau 97% dalam klasifikasi tweet ke dalam sentimen kelas negatif dan positif. Untuk akurasi masing – masing kelas sebesar

- a. Negatif

$$\frac{TN}{TN+FN} = \frac{48}{48+2} = \frac{48}{50} = 0,96$$

Hasil akurasi sentimen negatif mencapai 0,96 atau 96%, sehingga tingkat kesalahan klasifikasi yaitu 4%.

- b. Positif

$$\frac{TP}{TP+FP} = \frac{49}{49+1} = \frac{49}{50} = 0,98$$

Hasil akurasi sentimen positif mencapai 0,98 atau 98%, sehingga tingkat kesalahan klasifikasi kelas positif sebesar 2%.

5. KESIMPULAN

Dari hasil pengujian, dapat disimpulkan beberapa hal sebagai berikut :

- a. Metode Naive Bayes Classifier dalam melakukan klasifikasi tweet sentimen negatif dan positif dengan 300 data latih dan 100 data uji mendapat hasil akurasi sebesar 97%.
- b. Dengan hasil akurasi yang cukup tinggi yaitu 97% maka metode Naive Bayes Classifier dapat digunakan untuk melakukan klasifikasi tweet dengan sentimen negatif dan positif secara otomatis

6. SARAN

Dari hasil penelitian yang telah dilakukan, sistem yang dibangun masih memiliki kekurangan baik dari segi fungsionalitas maupun data yang dimiliki. Oleh karena itu agar didapat sistem yang lebih handal dan akurat perlu dilakukan pengembangan dan penyempurnaan lebih lanjut. Adapun saran agar sistem dapat berfungsi dengan lebih baik lagi yaitu :

- a. Melakukan perbandingan dengan algoritma klasifikasi lain, untuk mengetahui kinerja dari algoritma naive bayes classifier.
- b. Pengembangan ke dalam bahasa pemrograman lain.

DAFTAR PUSTAKA

- [1] Maulana, A. (2016) Twitter Rahasiakan Jumlah Pengguna di Indonesia, Website: <https://www.cnnindonesia.com/teknologi/20160322085045-185-118939/twitter-rahasiakan-jumlah-pengguna-di-indonesia>, diakses 20 April 2018.
- [2] Brogan, C. (2010) *Social Media 101 Tactic and Tips to Develop Your Business Online*, Wiley, Manhattan.
- [3] Dyarsa, P. S. (2015) Analisis Sentiment Pada Sosial Media Twitter Menggunakan Naive Bayes Classifier Terhadap Kata Kunci “Kurikulum 2013”, *Skripsi*, Program Studi Teknik Informasi, Universitas Dian Nuswantoro, Semarang.
- [4] Faradhillah, N. Y., Kusumawardani, R. P., & Hafidz, I. (2016) Eksperimen Sistem Klasifikasi Analisa Sentimen Twitter pada Akun Resmi Pemerintah Kota Surabaya Berbasis Pembelajaran Mesin. *Seminar Nasional Sistem Informasi Indonesia*, 1(3), pp.16-24.
- [5] Lorosae, T. A., Prakoso, B. D., Saifudin, S., & Kusrin, K. (2018) Analisis Sentimen Berdasarkan Opini Masyarakat pada Twitter Menggunakan Naive Bayes, *Seminar Nasional Teknologi Informasi & Multimedia Online*, 6(1), pp.1-10.
- [6] Hanggara, S., Akhriza, T. M., & Husni, M. (2017) Aplikasi Web Untuk Analisis Sentimen Pada Opini Produk Dengan Metode Naive Bayes Classifier, *Prosiding Seminar Nasional Inovasi dan Aplikasi Teknologi di Industri*, 3(1), pp.33-1.