

Analisis Sentimen Politik Berdasarkan Big Data dari Sosial Media Instagram

by John Doe

Submission date: 27-Nov-2021 07:35PM (UTC+0700)

Submission ID: 1713545742

File name: 17523227_paper_amin.pdf (872.53K)

Word count: 2469

Character count: 15205

1 Analisis Sentimen Politik Berdasarkan Big Data dari Sosial Media Instagram

Abstrak— Di dunia politik saat ini situs media sosial Instagram sering digunakan menjadi sarana promosi politik dan menimbulkan banyaknya komentar dari pengguna baik positif maupun negatif. Kinerja Presiden yang sedang berjalan saat ini pun tidak luput dari banyaknya komentar pengguna Instagram. Adanya masalah tersebut penelitian ini memiliki tujuan untuk menganalisis sentimen terkait kinerja politik Presiden Joko Widodo berdasarkan komentar yang ada pada media sosial Instagram. Proses analisis dilakukan dengan memanfaatkan algoritma Naïve Bayes Classifier (NBC) dan ekstraksi fitur *Tf-idf* dengan total jumlah data 1152 komentar yang didapatkan dari beberapa postingan mengenai Presiden Joko Widodo. Berdasarkan eksperimen yang dilakukan ditemukan bahwa algoritma NBC mampu melakukan analisis sentimen terkait komentar tentang politik dengan baik dibuktikan dengan adanya nilai *Accuracy* 83%, *Precision* 81%, *Recall* 84% dan hasil dari analisis sentimen terhadap data komentar menunjukkan bahwa sentimen positif berdasarkan kemunculan kata didominasi oleh kata sehat, bangga, mantap, presiden, dan keren. Sedangkan untuk kelas sentimen negatif kata-kata yang mendominasi adalah PCR, bumn, hutan, dan vaksin.

Keywords—*Analisis Sentimen, Joko Widodo, Instagram, Naïve Bayes Classifier*

I. PENDAHULUAN

Teknologi digital saat ini telah mengalami pertumbuhan yang cepat dan mendorong masyarakat untuk selalu mengikuti kemajuan tersebut, tidak hanya masyarakat perkotaan saja namun masyarakat desa pun dituntut agar dapat mengikuti zaman yang pesat ini. Adanya kemajuan pada aktivitas yang modern seperti saat ini mewajibkan masyarakat untuk bergerak secara fleksibel dalam memenuhi segala aspek kebutuhan. Dalam Era revolusi industri saat ini yang banyak memanfaatkan jaringan internet secara intens berdampak pada perubahan digitalisasi yang terjadi pada segala aspek kehidupan. Termasuk kedalam hal penyajian pesan dari individu kepada individu maupun dari individu kepada khalayak umum[1].

Media sosial merupakan sebuah tempat bagi individu untuk menuliskan sebuah gagasan, ide, maupun pendapat sehingga dapat diketahui berbagai individu lain. Contoh media sosial yang banyak dipakai masyarakat umum adalah media sosial Instagram. Aplikasi Instagram merupakan media sosial yang digunakan oleh banyak manusia di muka bumi salah satunya adalah negara Indonesia. Berdasarkan fungsinya Instagram digunakan untuk berbagi foto serta video pada sesama pengguna aplikasi tersebut. Dalam dunia politik Instagram sering digunakan menjadi sarana kampanye, promosi maupun untuk sosialisasi yang efisien dan efektif terhadap kebijakan yang berbau politik. Dalam postingan media sosial Instagram terdapat banyak komentar yang dilakukan oleh sesama pengguna sehingga dapat memunculkan opini dari masyarakat, berupa komentar positif maupun negatif[2].

Kinerja Presiden yang saat ini sedang berjalan juga tidak lepas dari komentar individu dalam media sosial Instagram. Presiden Joko Widodo mendapatkan banyak komentar, seperti

pujian, sebuah saran, kritikan hingga sebuah celaan. Dengan banyaknya antusias pengguna instagram dalam menyampaikan pendapat terkait kinerja presiden Joko Widodo dapat dimanfaatkan untuk mencari sebuah informasi sebagai bahan evaluasi kinerja politik presiden Joko Widodo. Untuk bisa memanfaatkannya membutuhkan analisis yang tepat sehingga segala informasi yang didapat bisa untuk mendukung suatu keputusan atau pilihan. Upaya yang dapat digunakan untuk menganalisis komentar yang ada pada instagram adalah salah satunya analisis sentimen[3]. Analisis sentimen atau bisa juga disebut *opinion mining* adalah sebuah proses pemahaman serta mengolah data yang berbentuk tekstual dengan secara otomatis untuk menghasilkan sebuah informasi berupa sentimen yang ada di satu kalimat pendapat. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan perangkat lunak berbasis analisis sentimen berkembang begitu pesat. Bahkan negara Amerika Serikat berkisar 20 hingga 30 perusahaan fokus terhadap analisis sentimen[4]. Pada penelitian kali ini analisis sentimen memiliki tujuan untuk melihat kecenderungan komentar kepada kinerja politik Presiden Joko Widodo mengandungi kelas sentimen positif atau negatif serta melihat kata apa saja yang mendominasi dalam sentimen positif maupun negatif sehingga dapat dimanfaatkan untuk evaluasi kebijakan publik kedepannya bagi pemerintahan Presiden Joko Widodo.

II. KAJIAN PUSTAKA

Analisis sentimen dengan menggunakan sumber big data dari media sosial bukan lagi sebuah hal yang baru terdapat beberapa penelitian sebelumnya yang pernah dilakukan. Berdasarkan penelitian Fadli, et al [5] analisis sentiment yang ada pada jejaring sosial Instagram dapat dilakukan dengan algoritma Decision Tree dan juga Support Vector Machine masing-masing memiliki akurasi yang terbilang sangat baik angka yang diperoleh sebesar 87,45% untuk Decision Tree dan 94,36% untuk SVM.

Pada penelitian yang dilakukan Mahardika, et al [6] Berfokus terhadap analisis sentimen kepada pemerintahan Joko Widodo pada ranah jejaring sosial twitter dalam penelitian tersebut data diperoleh dengan memakai API yang disediakan oleh sosial media yang dipakai. Pada penelitian ini menggunakan algoritma naive Bayes sebagai model analisis sentimen dengan performa terbaik yang didapat memiliki nilai akurasi sebesar 82,90% dan nilai $\alpha = 0,05$.

Penelitian tentang analisis sentimen pada media sosial Instagram menggunakan NBC sudah dilakukan sebelumnya oleh Hidayatullah, et al [7] Dari hasil eksperimen penelitian bahwa algoritma Naive Bayes memiliki performa amat baik yaitu akurasi didapatkan 85,27%, nilai presisi didapatkan 88,29% dan nilai *Recall* didapatkan 81,26%. Nilai akurasi paling tinggi diperoleh pada $k = 10$ dengan fold yang ke 10 yaitu 0,91. Nilai *precision* paling tinggi didapat pada $k = 10$ fold ke 10 yaitu 0,9 dan *Recall* paling tinggi diperoleh pada $k = 7$ fold ke 7 yaitu 0,91.

Ada pula penelitian analisis terkait isu politik para calon presiden dengan memakai metode SVM yang dilakukan oleh

Wahyu, et al [8] dalam penelitian tersebut menggunakan data dari media sosial twitter dengan hasil analisis memakai algoritma SVM terhadap sentimen Politik Calon Presiden 2019 ditemukan akurasi metode SVM kepada calon presiden Joko Widodo bernilai 98 %, serta akurasi kepada calon presiden Prabowo subianto bernilai 99%.

Bhumika, et al[9] melakukan perbandingan beberapa algoritma untuk melakukan analisis sentimen menggunakan data dari twitter dan dataset kaggle dengan hasil akurasi SVM : 85 %, Bayesian Logistic Regression: 74,84%, Naive Bayes: 66,24%, Random Forest Classifier: 87,5%, Neural Network: 89,93%. Penelitian mengenai analisis sentimen terlebih pada politik memang layak untuk diteliti lebih lanjut dan algoritma Naive Bayes Classifier (NBC) dipilih dalam penelitian ini karena memang terbukti memiliki akurasi yang sangat baik untuk membantu dalam proses analisis.

III. METODOLOGI PENELITIAN

Berikut merupakan beberapa alur yang dilaksanakan pada penelitian kali ini:

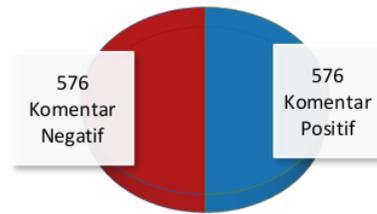


Gambar 1. Alur Penelitian

A. Dataset

Pengumpulan seluruh data diperoleh dari sumber data komentar media sosial Instagram yang membahas tentang politik kinerja Presiden Joko Widodo dengan batasan akun Instagram yang sudah *verified*. List akun Instagram yang dipilih pada penelitian ini adalah Detikcom, Kompascom, cnnindonesia, tempodotco, dan liputan6. Proses Scraping data komentar pada penelitian ini menggunakan Selenium serta menggunakan *Chrome Driver* sebagai perantaranya. Jumlah data keseluruhan yang diperoleh adalah 1152 komentar dengan rincian data berlabel positif berjumlah 576 dan data berlabel negatif berjumlah 576. Berikut merupakan representasi rincian *labelling* data dalam bentuk diagram ditunjukkan pada Gambar 2.

PELABELAN DATA



Gambar 2. Labelling Data

B. Pre-Processing

Setelah pengumpulan data selesai berikutnya adalah *Pre-Processing* yang berupa proses awalan yang dilakukan pada proses *text mining* bertujuan untuk menyiapkan data untuk memudahkan prediksi/klasifikasi. Berikut merupakan beberapa *Pre-Processing* yang dilaksanakan dalam penelitian ini:

- *Case Folding*
langkah proses membuat huruf yang ada pada data input menjadi keseluruhan huruf kecil.
- *Tokenizing*
Mengubah data menjadi beberapa bagian bagian yang lebih kecil (token) membuatnya lebih mudah untuk di analisa.
- *Filtering*
Mengubah beberapa kata imbuhan menjadi kata dasar aslinya serta menghapus kata yang kurang memiliki makna penting.
- *Stemming*
Digunakan mengambil semua teks yang bernilai penting yang bersumber dari token sebelumnya.
- *Spell Checker*
Digunakan untuk mengoreksi teks yang salah dalam proses mengetik/*typo* dan menggantinya dengan teks yang benar.

Berikut merupakan hasil *Pre-Processing* yang dilakukan pada data komentar Instagram ditunjukkan dalam Tabel 1.

Tabel 1. Hasil *Pre-Processing*

Sebelum	Sesudah
Mantap jokowi salam 3 periode	mantap jokowi salam periode
"Langsung ganti presiden aja , lebih cepat lebih baik buat bangsa dan rakyat Indonesia !!!!"	langsung ganti presiden aja cepat bangsa rakyat indonesia
Bikin malu...nanti pemimpin negara sahabat tanya kenapa pesawat kepresidenan gak dipake? mau jawab apa hayoooo	bikin malu pimpin negara sahabat pesawat presiden gak dipake hayoooo

C. Ekstraksi Fitur

Data yang telah melalui *Pre-Processing* berupa *text* akan diubah kedalam bentuk numerik dan dilakukan pembobotan kata yang memiliki tujuan untuk melakukan perhitungan bobot pada setiap kata yang nantinya akan digunakan sebagai fitur. Pada tahapan ini terdapat dua buah proses yaitu TF dan IDF, TF adalah munculnya kata pada setiap data semakin banyak jumlah kata yang muncul maka akan semakin besar juga nilai dari TF nya. IDF merupakan nilai yang ada pada tiap kata dan berbanding terbalik sehingga apabila kata yang ada jarang ditemui pada data maka IDF akan memiliki nilai lebih tinggi [10]. Berikut merupakan rumus dari *Tf-idf* ditunjukkan oleh rumus nomor (1).

$$W_{i,j} = tf_{i,j} \cdot \log\left(\frac{N}{df_i}\right) \quad (1)$$

Keterangan:

$W_{i,j}$ = Bobot *Tf-idf*

$tf_{i,j}$ = Seluruh frekuensi data

N = Seluruh total Dokumen

df_i = Seluruh frekuensi dokumen tiap data

D. Klasifikasi

Metode klasifikasi yang digunakan untuk analisis sentimen dan mendapatkan prediksi positif atau negatif pada penelitian kali ini adalah *Naïve Bayes Classifier*. *Naïve Bayes* adalah sebuah metode klasifikasi yang dihitung dari kombinasi probabilitas jumlah dan nilai frekuensi dari data yang ada. Pada pembuatan model klasifikasi ini menggunakan data latihan/*training* sebanyak 80% dan 20% sebagai data uji/*testing* dari jumlah data yang telah dilakukan *Pre-Processing* sebelumnya. Berikut merupakan persamaan *Naïve Bayes* ditunjukkan oleh rumus nomor (2).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2)$$

Keterangan:

$P(H|X)$: Peluang nilai hipotesa H dengan kondisi X

$P(H)$: Peluang hipotesa H

$P(X|H)$: Peluang X dengan kondisi hipotesa H

$P(X)$: Peluang hipotesa X

Naïve Bayes Classifier mempunyai beberapa aturan yaitu jika $P(H_q|x) < P(H_r|x)$, maka nilai x akan dideteksi H_r [11]. Pernyataan $P(H_q|x)$ memiliki indikasi peluang hipotesis H_q dengan berdasarkan kondisi x, berlaku juga dengan H_r .

E. Evaluasi

Setelah dilakukan proses klasifikasi menggunakan algoritma *Naïve Bayes Classifier* (NBC) selanjutnya adalah proses evaluasi yang bertujuan untuk melihat seberapa baik model klasifikasi yang dihasilkan. Proses evaluasi dilakukan menggunakan tabel *Confusion Matrix* yang ditunjukkan Tabel 3.

Tabel 2. *Confusion Matrix*

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	TP	FN

Negatif	FP	TN
---------	----	----

Serta nilai dalam *Confusion Matrix* dapat diselesaikan dengan menghitung nilai *Accuration*, *Precision*, dan *Recall* [12] didasarkan pada rumus nomor (3), (4) dan (5).

$$Accuration = \frac{TP + TN}{TP + TN + FN + FP} \quad (3)$$

$$Precision = \frac{TP}{FP + TP} \quad (4)$$

$$Recall = \frac{TP}{FN + TP} \quad (5)$$

Keterangan:

TP (*true positive*): dokumen yang nilainya positif dan diestimasi benar sebagai positif.

FP (*false positive*): dokumen yang nilainya negatif dan diestimasi salah sebagai positif.

TN (*true negative*): dokumen yang nilainya negatif dan diestimasi benar sebagai negatif.

FN (*false negative*): dokumen yang nilainya positif dan diestimasi salah sebagai negatif.

IV. HASIL DAN PEMBAHASAN

A. Perhitungan *Accuracy*, *Precision*, dan *Recall*

Berikut merupakan hasil *Confusion Matrix* pada data *testing* ditunjukkan pada Tabel 3.

Tabel 3. Hasil *Confusion Matrix*

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	91	20
Negatif	17	103

Pada Tabel 3. Menunjukkan bahwa model yang telah dibangun dengan *Naïve Bayes Classifier* (NBC) dan ekstraksi fitur *Tf-idf* dapat memprediksi sentimen positif dengan benar sebanyak 91 komentar dari seluruh jumlah data. Dari nilai yang ada pada tabel diatas dapat dihitung *Accuracy* menggunakan rumus nomor (3), *Precision* menggunakan rumus nomor (4), dan *Recall* menggunakan rumus nomor (5). Berikut merupakan hasil perhitungannya ditunjukkan oleh Tabel 4.

Tabel 4. *Accuracy*, *Precision*, dan *Recall*

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
0,83	0,81	0,84

Hasil yang diperlihatkan pada Tabel 4. Dapat dituliskan nilai *Accuracy*, *Precision*, dan *Recall* yang didapatkan dari hasil uji model menggunakan data *testing* menghasilkan nilai yang sangat baik yaitu nilai *Accuracy* sebesar 0,83, nilai *Precision* sebesar 0,81, dan nilai *Recall* 0,84.

B. Analisis Kata Berdasarkan Sentimen

Penelitian kali ini juga melakukan berupa analisis data yang berisi kemunculan kata yang paling sering muncul pada

data komentar pengguna Instagram terkait kinerja politik Presiden Joko Widodo. Kemunculan kata dipecah menjadi dua buah kelas yaitu positif dan negatif. Berikut merupakan himpunan kata yang sering dibahas ditampilkan dengan *Wordcloud* yang diperlihatkan pada Gambar 3. dan Gambar 4.



Gambar 3. *Wordcloud* Positif



Gambar 4. *Wordcloud* Negatif

Pada gambar diatas kelas *Wordcloud* yang memiliki sentimen positif kemunculan kata yang ada didominasi kata sehat, bangga, mantap, presiden, dan keren. Sedangkan kelas sentimen negatif kata-kata yang mendominasi adalah PCR, bumn, hutan, dan vaksin.

V. KESIMPULAN

Hasil dari seluruh alur penelitian terkait analisis sentimen kinerja politik Presiden Joko Widodo dengan menggunakan algoritma Naïve Bayes Classifier (NBC) diperoleh beberapa kesimpulan yaitu:

1. Algoritma Naïve Bayes Classifier terbukti mampu melakukan analisis sentimen dengan sangat baik dibuktikan dengan Nilai *Accuracy* sebesar 83%,

Nilai *Precision* bernilai 81% dan Nilai *Recall* sebesar 84%.

2. Dari hasil analisis kata terhadap data yang ada menunjukkan kata sehat, bangga, mantap, presiden, dan keren mendominasi kelas sentimen positif dan kata PCR, bumn, hutan, dan Indonesia mendominasi kelas sentimen negatif.
3. Penelitian ini masih terdapat keterbatasan mengklasifikasikan komentar hanya kedalam kelas positif dan negatif. Penulis berharap penelitian ini dapat dikembangkan lagi dengan menambahkan fitur klasifikasi yang lebih spesifik seperti komentar dengan kelas kritik, saran, hujatan, pengancaman dan lain sebagainya serta menambahkan jumlah data *training* agar model yang dihasilkan semakin baik.

REFERENSI

- [1] I. Kurniasari and H. Al Fatta, "Analisis Sentimen Opini Publik pada Instagram mengenai Covid-19 dengan SVM," *JTECS J. Sist. Telekomun. Elektron. Sist. Kontrol Power Sist. Komput.*, vol. 1, no. 1, pp. 67–74, 2021.
- [2] N. D. Amin, "Klasifikasi Konten Instagram Berdasarkan Komentar Menggunakan Support Vector Machine," 2019.
- [3] R. A. Wildan, R. A. Rajagede, and R. Rahmadi, "Analisis Sentimen Politik Berdasarkan Big Data dari Media Sosial Youtube : Sebuah Tinjauan Literatur," *Automata*, vol. 2, 2021.
- [4] G. A. Buntoro, "Analisis Sentimen Calon Gubemur DKI Jakarta 2017 Di Twitter," *INTEGER J. Inf. Technol.*, vol. 1, no. 1, pp. 32–41, 2017, [Online]. Available: https://www.researchgate.net/profile/Ghulam_Buntoro/publication/n/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf.
- [5] K. M. L. Muhammad Fadli Ashiddiqi, "Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI," vol. 1, no. 3, pp. 71–75, 2020.
- [6] Y. S. Mahardika and E. Zuliarso, "Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes," *Pros. SINTAK 2018*, no. 2015, pp. 409–413, 2018.
- [7] R. Hidayatullah, U. A. R. M. Kom, G. Abdurrahman, S. Si, and M. Pd, "Produk Eiger Di Instagram Menggunakan Metode Naive Bayes Classifier," pp. 1–8.
- [8] W. P. Jati, A. Z. Lutfi, H. Ilma, and E. Widodo, "Analisis Sentimen terhadap Isu Politik Calon Presiden Indonesia 2019 berdasarkan Opini Netizen dari Twitter menggunakan Metode Klasifikasi Support Vector Machine," *Pros. Konf. Nas. Penelit. Mat. dan Pembelajarannya IV 2019*, no. Prodi Pendidikan Matematika FKIP UMS, p. PROSIDING--M2, 2019.
- [9] B. Gupta, M. Negi, K. Vishwakarma, G. Rawat, and P. Badhani, "Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python," *Int. J. Comput. Appl.*, vol. 165, no. 9, pp. 29–34, 2017, doi: 10.5120/ijca2017914022.
- [10] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest

Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.

- [11] Z. Azmiardhy, F. Rachma, and A. F. Hidayatullah, "Deteksi Surel Spam dan Non Spam Bahasa Indonesia Menggunakan Metode Naïve Bayes," 2021.

[12] N. Novalita, A. Herdiani, and I. Lukmana, "Identifikasi Cyberbullying Pada Media Sosial Twitter Menggunakan Metode Klasifikasi Random Forest," 2019.

Analisis Sentimen Politik Berdasarkan Big Data dari Sosial Media Instagram

ORIGINALITY REPORT

12%

SIMILARITY INDEX

11%

INTERNET SOURCES

7%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1	123dok.com Internet Source	2%
2	repository.unmuhjember.ac.id Internet Source	1%
3	www.researchgate.net Internet Source	1%
4	jurnal.stts.edu Internet Source	1%
5	repository.uinjkt.ac.id Internet Source	1%
6	docplayer.info Internet Source	1%
7	Yudo Bismo, Giofani Harsanto. "Penerapan Metode Certainty Factor Dan Naïve Bayes Untuk Mendiagnosa Penyakit Akibat Gigitan Nyamuk", Generation Journal, 2020 Publication	1%
8	sipil.studentjournal.ub.ac.id Internet Source	

1 %

9

digilib.uin-suka.ac.id

Internet Source

<1 %

10

dokumen.pub

Internet Source

<1 %

11

Caroline Layadi, Mohammad Fajar, Hasniati Hasniati, Izmy Alwiah Musdar. "Analisis Data Pada Jaringan Sensor Nirkabel Menggunakan Metode Support Vector Machine", Jurnal RESISTOR (Rekayasa Sistem Komputer), 2018

Publication

<1 %

12

cdn.repository.uisi.ac.id

Internet Source

<1 %

13

dspace.uii.ac.id

Internet Source

<1 %

14

ejournal.uin-suska.ac.id

Internet Source

<1 %

15

repository.usd.ac.id

Internet Source

<1 %

16

Euis Saraswati, Yuyun Umaidah, Apriade Voutama. "Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap Covid-19", Generation Journal, 2021

Publication

<1 %

17

Riska Aryanti, Atang Saepudin, Eka Fitriani, Rifky Permana, Dede Firmansyah Saefudin. "Komparasi Algoritma Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Pengguna Busway", Jurnal Teknik Komputer, 2019

Publication

<1 %

18

download.garuda.ristekdikti.go.id

Internet Source

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On