

Analisis Sentimen Terhadap Tokoh Publik Menggunakan Support Vector Machine

¹⁾ Nurul Khasanah Fitriyani

Universitas Amikom Yogyakarta, Jl. Ringroad Utara, Yogyakarta, Indonesia
E-Mail: nurulkhafitriyani17@gmail.com

²⁾ Anggit Dwi Hartanto

Universitas Amikom Yogyakarta, Jl. Ringroad Utara, Yogyakarta, Indonesia
E-Mail: anggit@amikom.ac.id

ABSTRACT

Anies Baswedan was a governor of DKI Jakarta at the time of 2017-2022. In December 2019 It was warmly talked about due to the awarding of the award-winning Adikarya 2019 to the Colosseum discotheque although the award was eventually revoked. His name is also warmly talked about because it is considered washing hands after uninstalling two officials because of two different problems. Then also about the problem of flooding that occurred in Jakarta, the name is also called can not handle well the flood that always occurred in Jakarta. Twitter Media has a simple view, updated topics, open in accessing tweets and quickly in delivering opinions. Of the various comments and responses on Twitter required techniques to divide into negative or positive opinion classes. This research, using preprocessing and label opinions into positive and negative classes. As for the classifications using the Support Vector Machine method. The Data used is an opinion about the Anies Baswedan from social media Twitter which amounted to 1000 tweets taken on December 17, 2019. From the labeling result, many positive comments amounted to 429 and the negative comment amounted to 530. While the classification of the Support Vector Machine method is getting an accuracy value of 95.9%, the value of the presisis is 94.49% and the recall rate of 96.4%.

Keyword : Sentiment Analysis, Support Vector Machine, Twitter

PENDAHULUAN

Anies Baswedan adalah seorang gubernur DKI Jakarta yang menjabat pada masa bakti 2017-2022. Pada bulan Desember 2019 ini nama Anies Baswedan hangat diperbincangkan di berbagai media karena pemberian penghargaan Adikarya Wisata 2019 sebuah diskotek yaitu ke Diskotek Colosseum meskipun pada akhirnya penghargaan tersebut dicabut kembali. Namanya juga hangat diperbincangkan karena dianggap cuci tangan setelah mencopot dua pejabat karena dua masalah yang berbeda. Kemudian juga mengenai masalah banjir yang terjadi di Jakarta, namanya juga disebut belum bisa menangani dengan baik banjir yang selalu terjadi di Jakarta.

Twitter merupakan media sosial yang memiliki pengguna perharinya 126 juta. Selain itu twitter memiliki Twitter Trending Topic yang merupakan daftar topik terpopuler dan diupdate setiap waktu. Selanjutnya kelebihan lain dari twitter semua postingan yang ada di Twitter dapat diakses oleh pengguna lainnya. Dengan kemudahan tersebut pengguna Twitter dapat saling berinteraksi tanpa harus menjadi follower.

Dari penjelasan diatas twitter memiliki kelebihan dibanding media sosial lain yaitu tampilannya simple, topik terupdate, cepat dalam mengabarkan suatu opini pengguna dan terbuka

untuk mengakses tweet. Hal ini dapat semakin memudahkan dalam pengumpulan data opini dari masyarakat dan sebagai media penelitian. Agar mengetahui bagaimana pendapat masyarakat mengenai Anies Baswedan yang merupakan seorang gubernur DKI Jakarta diperlukan analisis sentimen terhadap twitter. Analisis sentimen dapat mengelompokkan opini untuk mengetahui sentimen positif atau sentimen negatif. Pada penelitian ini, akan dibahas tahapan yang dilalui untuk melakukan proses analisis sentimen. Dimulai dari tahapan pengumpulan data, preprocessing, kemudian tahapan analisis sentimen dengan menggunakan metode Support Vector Machine.

Menurut [1] melakukan sentimen analisis review barang berbahasa Indonesia di website dengan menggunakan metode SVM dan Query Expansion, dengan nilai akurasi 94,75%. Penelitian [2] melakukan sentimen analisis terhadap tingkat kepuasan pengguna marketplace online di Twitter menggunakan metode Naive Bayes Classifier dengan akurasi 87%. Analisis sentimen juga dilakukan oleh [3] menggunakan metode SVM. Sedangkan [4] menggunakan pendekatan Lexicon Based. Penelitian [5] melakukan perbandingan akurasi dan waktu proses pada algoritma KNN dan SVM untuk

analisis sentimen pada media sosial Twitter dan hasilnya metode SVM memiliki akurasi lebih unggul daripada metode KNN tetapi pada perhitungan waktu proses metode KNN lebih unggul daripada metode SVM. Pada penelitian [6] juga membandingkan kinerja algoritma C.45 dan Naive Bayes, hasilnya algoritma C45 memiliki akurasi yang lebih tinggi dibandingkan algoritma Naive Bayes. Analisis sentimen juga dilakukan oleh [7] untuk evaluasi faktor usability pada pengguna aplikasi mobile dengan model PACMAD (People At The Centre of Mobile Application Development) menggunakan metode TF-ICF dan kemudian dianalisis menggunakan metode SentiWordNet Interpretation. Sedangkan penelitian [8] menggunakan algoritma Naive Bayes dengan optimasi Feature Selection (FS) Particle Swarm Optimization menghasilkan peningkatan akurasi dan AUC yang signifikan.

BAHAN DAN METODE

Data yang digunakan dalam penelitian ini adalah data tweet dari media sosial Twitter. Data diambil menggunakan metode crawling dengan Rstudio. Sebelum proses crawling dilakukan, diperlukan API Key dari Twitter sebagai autentikasi dengan Rstudio. Kemudian diperlukan package untuk crawling twitter di Rstudio. Package diinstall dan proses crawling data twitter dapat dijalankan.

Preprocessing data tweet yang sebelumnya dari data yang tidak terstruktur menjadi data yang terstruktur sehingga data dapat dianalisis. Tahap preprocessing yang digunakan dalam penelitian ini adalah case folding, tokenizing, filtering dan stemming.

Selanjutnya, setelah data melewati tahap preprocessing kemudian data tweet akan ditentukan kedalam kelas opini yang dimunculkan dalam penelitian ini ada dua yaitu positif dan negatif.

Setelah pelabelan maka data tweet yang sudah dibagi kelasnya selanjutnya akan masuk ke tahap klasifikasi dengan menghitung akurasi algoritma SVM. Proses klasifikasi ini dibantu dengan Weka 3.8.

Selanjutnya untuk proses pengujian untuk melihat tingkat akurasi dan kesalahan dari sistem tersebut. Pengukuran kinerja klasifikasi dilakukan dengan confusion matriks. Confusion Matrix adalah tabel pencatat hasil kerja dari klasifikasi.

Tabel 1. Confusion Matriks

	Relevant	Not Relevant
Retrieved	True Positive	False Positive
Not Retrieved	False Negative	True Negative

Berdasarkan isi confusion matrix, dapat diketahui nilai dari presisi, recall dan akurasi dari hasil klasifikasi yang diperoleh. Untuk menghitung akurasi, presisi serta recall menggunakan formula

sebagai berikut:

$$\begin{aligned} \text{Presisi} &= TP / (TP+FP) \\ \text{Recall} &= TP / (TP +FN) \\ \text{Akurasi} &= (TP +TN) / (TD) \end{aligned}$$

Keterangan :

TP = True Positif / Jumlah Data Positif yang benar
 TN = True Negatif / Jumlah data Negatif yang Benar
 FN= False Negatif / Jumlah data Negatif Yang salah
 TD = Total data/ Seluruh Data yang diuji

HASIL DAN PEMBAHASAN

Pada tahap pengumpulan data twitter proses yang perlu dilakukan terlebih dahulu adalah mendapatkan API Key Twitter yang bisa didapatkan di <https://developer.twitter.com/apps>. Tahap pertama membuat aplikasi pada App Twitter, lalu mengisi form pengajuan API Key. Setelah berhasil mendaftar maka akan mendapatkan API Key yang berguna menghubungkan RStudio dengan Twitter.

Setelah mendapat API Key proses crawling dapat dilakukan. Pertama memanggil library di R Studio yang sudah diinstall yaitu "twitter", "RCurl", "tm", "corpus", "wordcloud".

Setelah autentikasi API Key berhasil, proses crawling dapat dilakukan dengan mencari kata "Anies Baswedan", data yang diambil berjumlah 1000. Data tweet diambil pada tanggal 17 Desember 2019.

Data telah diperoleh masih berupa data mentah dimana data tersebut masih banyak mengandung noise dan belum terstruktur. Oleh karena itu, perlu dilakukan preprocessing untuk menghilangkan noise dalam kata, mengkonversi data agar lebih terstruktur dan untuk mempermudah dalam penelitian. Pada penelitian yang akan dilakukan, tahap preprocessing yang akan digunakan adalah cleansing, case folding, filtering, dan tokenizing.

a. Cleansing

Cleansing merupakan membersihkan data tweet dari noise yang berupa mention username (@username), HTML, kata kunci, RT atau Retweet, hashtag (#), email, link tautan (<http://website.com>), dan mention (@username).

Tabel 2. Proses Cleansing

Input	Output
RT @bangzul_88: Setiap kebijakan Anies Baswedan yang menuai pro dan kontra dan dinilai tidak berpihak masyarakat. Sikap Anies adalah cuci tangan dan tak mau	Setiap kebijakan Anies Baswedan yang menuai pro dan kontra dan dinilai tidak berpihak masyarakat Sikap Anies adalah cuci tangan dan tak mau

disalahkan, #4niesCuciTangan	disalahkan
Posisi yang pantas cocok buat Anies Baswedan sebetulnya adalah Jubir... bukan pemimpin... https://t.co/zkKzsrgY0n	Posisi yang pantas cocok buat Anies Baswedan sebetulnya adalah Jubir bukan pemimpin

b. Case Folding

Case folding merupakan proses mengubah huruf kapital menjadi huruf kecil (lowercase). Proses ini diperlukan karena huruf kapital tidak konsisten dalam penggunaannya.

Tabel 3. Proses Case Folding

Input	Output
Setiap kebijakan Anies Baswedan yang menuai pro dan kontra dan dinilai tidak berpihak masyarakat Sikap Anies adalah cuci tangan dan tak mau disalahkan	setiap kebijakan anies baswedan yang menuai pro dan kontra dan dinilai tidak berpihak masyarakat sikap anies adalah cuci tangan dan tak mau disalahkan
Posisi yang pantas cocok buat Anies Baswedan sebetulnya adalah Jubir bukan pemimpin	posisi yang pantas cocok buat anies baswedan sebetulnya adalah jubir bukan pemimpin

c. Filtering

Tahap Filtering merupakan tahap penyaringan kalimat tweet atau pengurangan kata yang didalam corpus yang disebut stopword. Stopword sendiri adalah menghilangkan kata yang tidak berpengaruh atau tidak memiliki makna serta menghilangkan kata kunci dari proses crawling dan kata yang tidak diinginkan.

Tabel 4 Proses Filtering

Input	Output
Setiap kebijakan anies baswedan yang menuai pro dan kontra dan dinilai tidak berpihak masyarakat sikap anies adalah cuci tangan dan tak mau disalahkan	kebijakan menuai pro kontra dinilai berpihak masyarakat sikap tangan disalahkan
Posisi yang pantas cocok buat anies baswedan sebetulnya adalah jubir bukan pemimpin	posisi pantas jubir pemimpin

d. Tokenizing

Tahap terakhir dalam preprocessing ini adalah tokenizing. Tokenizing sendiri adalah proses

memecah kata per kata pada sebuah kalimat atau dokumen. Proses ini diperlukan karena untuk mendapatkan kata yang memiliki nilai. Tokenizing ini dapat mempermudah menghitung frekuensi kemunculan kata dalam corpus.

Tabel 5 Proses Tokenizing

Input	Output
kebijakan menuai pro kontra dinilai berpihak masyarakat sikap cuci tangan disalahkan	kebijakan menuai pro kontra dinilai berpihak masyarakat sikap cuci tangan disalahkan
Posisi yang pantas cocok buat anies baswedan sebetulnya adalah jubir bukan pemimpin	posisi pantas cocok jubir pemimpin

Setelah preprocessing selesai maka tahap selanjutnya adalah membangun Term Document Matrix (TDM), TDM sendiri adalah bentuk matriks matematika yang menggambarkan frekuensi kata yang sering muncul dalam corpus, dalam penelitian ini menampilkan frekuensi kata dari 1000 tweet menghasilkan 712 kata.

word	freq
dinilai	373
niescucitangan	349
kebijakan	296
pro	290
kontra	288
menuai	288
rakyat	287
penghargaan	267
dipecat	145
foto	144
opo	98
korban	91

Showing 1 to 12 of 712 entries

Gambar 1. Term Document Matrix (TDM)

Pelabelan Kelas Opini

Pelabelan kelas opini atau sentimen dimana memberikan label pada kalimat atau dokumen ke dalam kelas positif atau negatif. Hasil kalimat dari perhitungan skor yang >0 akan dilabeli kelas positif, begitupun sebaliknya jika hasilnya <0 maka dilabeli negatif.

score	text	klasifikasi
1	bang pejabat plt kadis persetujuan izin diberhentikan	Positif
2	-1 rakyat niescutitangan kebijakan menuai pro kontra dinil...	Negatif
3	-1 kasihan deh anggotamu om akibat dr arogansimu angg...	Negatif
4	0 tunggu lord menyalahkan minimal nih ujan disalahin da...	Positif
5	-1 buzzer kodian penjilat sekedar point kritik tp menginga...	Negatif
6	-1 rakyat niescutitangan kebijakan menuai pro kontra dinil...	Negatif
7	0 air dialirkan kelaut menyalahi takdir allah niescutitangan	Positif
8	-2 salah perubahan yg foto foto sungai citarum administrat	Negatif
9	0 banjir dki mendengarkan teori	Positif
10	0 banjir dki mendengarkan teori	Positif
11	0 banjir dki mendengarkan teori	Positif
12	1 alhamdulillah pemimpin mendengar aspirasi keluhan w...	Positif

Gambar 2. Pelabelan Kelas Opini

Klasifikasi

Dalam penelitian ini data diuji menggunakan 10 fold cross validation, dimana data dibagi menjadi dua yaitu 9/10 bagian adalah data training dan sisanya 1/10 bagian adalah data testing. Pengujian dilakukan sampai 10 kali dengan kombinasi yang berurutan.

```

Time taken to build model: 0.67 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      959      95.9 %
Incorrectly Classified Instances    41       4.1 %
Kappa statistic                    0.9172
Mean absolute error                 0.041
Root mean squared error             0.2025
Relative absolute error             8.3001 %
Root relative squared error        40.7437 %
Total Number of Instances         1000

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.  0,959  0,040  0,959  0,959  0,959  0,917  0,959  0,941

=== Confusion Matrix ===
  a  b  <- classified as
429 16 | a = Positif
 25 530 | b = Negatif
    
```

Gambar 3. Klasifikasi SVM

Dari hasil penelitian diatas dilakukan evaluasi hasil untuk mengukur dan dapat menarik kesimpulan dari proses penelitian yang telah dilakukan. Pengukuran ini dilakukan untuk menggambarkan seberapa baik sistem dalam proses mengklasifikasikan data. Dalam penelitian sebelumnya banyak menggunakan metode untuk evaluasi hasil salah satunya Confussion Matrix. Dari Confussion Matrix akan dihitung akurasi, presisi, dan recall.

Tabel 5. Confussion Matriks

	Positif	Negatif
Positif	429	16
Negatif	25	530

Dari data Confussion Matrix maka dapat dihitung akurasi, presisi, dan recall sebagai berikut.

$$\text{Akurasi} = \frac{429 + 530}{429 + 16 + 25 + 530} \times 100\% = 95,9\%$$

$$\text{Presisi} = \frac{429}{429 + 25} \times 100\% = 94,49\%$$

$$\text{Recall} = \frac{429}{429 + 16} \times 100\% = 96,4\%$$

Dari hasil diatas yang didapatkan adalah akurasi sebesar 95,9%, presisi sebesar 94,49%, dan recall sebesar 96,4%.

KESIMPULAN

Berdasarkan hasil dari penelitian yang telah dilakukan, maka dapat disimpulkan bahwa.

1. Dari penelitian ini digunakan untuk mengetahui ulasan komentar di media sosial Twitter didapatkan banyaknya berkomentar positif berjumlah 429 dan pengguna yang berkomentar negatif berjumlah 530.
2. Dari penelitian ini juga menunjukkan metode klasifikasi SVM mendapatkan akurasi sebesar 95,9% dimana sistem kernel dapat melakukan klasifikasi dengan sebanyak 959 ulasan.
3. Dari penelitian ini, komentar negatif lebih tinggi daripada komentar positif.

DAFTAR PUSTAKA

- [1] D. J. Haryanto, L. Muflikhah, and M. A. Fauzi. 2018. "Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Bravijaya*, vol. 2, no. 9, pp. 2909–2916.
- [2] H. A. F. Darmanto, Mohammad Suyanto. 2018. "Sentimen Analisis Review Pengguna Marketplace Online Menggunakan Naïve Bayes Classifier," *Pemodelan Arsit. Sist. Inf. Perizinan Menggunakan Kerangka Kerja Togaf Adm*, vol. 3, no. 1.
- [3] R. N. Chory, M. Nasrun, and C. Setianingsih. 2019. "Sentiment analysis on user satisfaction level of mobile data services using Support Vector Machine (SVM) algorithm," *Proc. - 2018 IEEE Int. Conf. Internet Things Intell. Syst. IOTAIS 2018*, pp. 194–200. doi: 10.1109/IOTAIS.2018.8600884.
- [4] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan. 2019. "Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine," *Komputek*, vol. 3, no. 2, p. 52. doi: 10.24269/jkt.v3i2.270.
- [5] M. R. A. Nasution and M. Hayaty. 2019. "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 226–235, doi: 10.31311/ji.v6i2.5129.
- [6] N. Yahya et al. 2019. "Komparasi Kinerja Algoritma C.45 Dan Naive Bayes Untuk Prediksi Kegiatan Penerimaan mahasiswa Baru (Studi Kasus : Universitas Stikubank Semarang)," *Prosiding SENDI*, no. 2014, pp. 978–979.
- [7] S. R. Wardhana, D. Purwitasari, and S. Rochimah. 2016. "Analisis Sentimen Pada

Review Pengguna Aplikasi Mobile Untuk Evaluasi Faktor Usability," *J. Sist. dan Inform.*, vol. 11, no. 1, pp. 128–136.

- [8] R. Sistem, S. A. Saputra, D. Rosiyadi, W. Gata, and S. M. Husain. 2021. "Analisis

Sentimen E-Wallet Pada Google Play Menggunakan Algoritma," vol. 1, no. 10, pp. 3–8.

