ANALISIS SENTIMEN CALON PRESIDEN DAN WAKIL PRESIDEN PERIODE 2019-2024 PASCA DEBAT PILPRES DI TWITTER

Sentiment Analysis of Presidential and Vice Presidential Candidates for the Period 2019-2024 After the debate on the Presidential Election on Twitter

Rizka Ardiansyah

Jurusan Teknologi Informasi, Fakultas Teknik, Universitas Tadulako e-mail: rizka@untad.ac.id / ardiansyah.rizka@gmail.com

Abstract

Social networking sites, such as Twitter and Facebook are one of the important spaces for political engagement. Twitter or Facebook have become common elements in political campaigns and elections, especially for Indonesia's presidential election 2019, for the period 2019 - 2024 there are two presidential and vice presidential candidates namely Ir. H. Joko Widodo - Prof. Dr. K.H. Ma'ruf Amin and Lieutenant General (Ret.) H. Prabowo Subianto - H. Sandiaga Uno. B.B.A., M.B.A. the two candidates who ran for the election triggered a lot of related public opinion where the most suitable candidate to become the president of the next period. Public opinion is generally one of the determining factors for presidential candidates who will later win the election. Presidential candidate debate is the efforts of the election commission to facilitate the presidential candidates to introduce their work programs to the public while building public opinion that they are the right people to become leaders of the next period. Although of course, this is not the only major factor that shapes public opinion. The purpose of this study is to summarize the opinions of the people voiced through social media related to the election of candidates for the Indonesian President and Vice President for the period 2019-2024 post debate on the presidential election. While the benefit is to help the community so that they can understand in a broader context such as what the public opinion about presidential candidates, especially on social media Twitter. The results of this study were presidential candidate Joko Widodo - Makruf Amin obtained a 25% positive sentiment, 4.5% negative sentiment and 70.5% neutral sentiment. while the Prabowo Subianto - Sandiaga Uno pair received a 5.1% positive sentiment, 2.5% negative sentiment and 92.4% neutral sentiment.

Keywords—Sentiment Analysis, 2019 Indonesian President Elections, Naïve Bayes Clasifier

1. Pendahuluan

Tahun 2019 merupakan tahun pesta demokrasi terbesar bagi rakyat republik Indonesia karena pada tahun ini masa jabatan Presiden dan Wakil Presiden Republik Indonesia Periode 2014-2019 berakhir. Kemeriahan dan antusiasme masyarakat terkait pasangan yang mencalonkan diri tidak hanya nampak di dunia nyata, namun juga di dunia maya melalui berbagai situs jejaring sosial seperti Facebook dan Twitter. Tahapan pemilihan presiden tahun 2019 sudah diumumkan oleh Komisi Pemilihan Umum (KPU) melalui portal resminya sejak tahun sebelumnya [1]. Sejak masih tahap pendaftaran sampai ditetapkannya siapa calon Presiden dan Wakil Presiden periode 2019 - 2024 yang bakal maju di PEMILU Indonesia 2019, nama-nama calonnya sudah mulai banyak diperbincangkan. Dunia maya yang begitu bebas dan sulit dikontrol, membuat semua orang bebas

berpendapat atau beropini tentang calon Presiden dan Wakil Presiden jagoannya masing-masing, memunculkan banyak opini-opini masyarakat, tidak hanya opini yang positif atau netral tapi juga yang negatif.

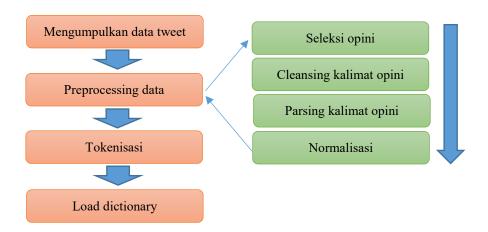
Untuk periode 2019-2024 ada dua calon presiden dan wakil presiden yang telah resmi dan lolos persyaratan untuk mencalonkan diri yaitu Ir.H.Joko Widodo - Prof.Dr.K.H. Ma'ruf Amin dan Letnan Jendral (Purn) H. Prabowo Subianto – H. Sandiaga Uno. B.B.A., M.B.A. Debat kandidat Presiden atau yang umum di sebut Debat CAPRES – CAWAPRES 2019 yang telah disiarkan dalam lima babak di berbagai stasiun televisi nasional merupakan upaya Komisi Pemilihan Umum (KPU) untuk menfasilitasi kandidat presiden dan wakil presiden untuk memperkenalkan program kerja kepada publik sambil membangun opini publik bahwa mereka adalah orang yang tepat untuk menjadi Presiden dan Wakil Presiden periode berikutnya [2]. Selain itu agar PASLON dan program kerjanya dapat lebih dikenal oleh masyarakat luas, PASLON tersebut juga melakukan kampanye politik. Selain menggunakan media konvensional, media elektronik seperti media sosial khususnya Twitter adalah salah satu media favorit PASLON dan tim suksesnya untuk melakukan kampanye karena dinilai lebih efektif dan efisien. Hal ini memicu berbagai opini di Twitter. Opini publik memiliki peran penting dalam menyukseskan PASLON dalam pemilu, seperti yang disampaikan oleh Freddy [3], bahwa opini publik dapat memberikan pengaruh kepada orang lain dalam menentukan sikap. Opini ini dapat dimanfaatkan untuk melihat bagaimana polaritas suatu tokoh politik yang akan maju sebagai calon Presiden dan Wakil Presiden Indonesia tahun 2019. Penentukan polaritas positif atau negatifnya suatu opini dapat dilakukan secara manual, tetapi seiring bertambahnya sumber opini yang semakin banyak, tentunya waktu dan usaha yang dibutuhkan untuk mengklasifikasikan polaritas opini tersebut akan semakin banyak [4]. Besarnya polaritas yang ditujukan kepada suatu paslon bisa dijadikan sebuah parameter kemenangan atau kekalahan paslon tersebut [5].

Pada penelitian ini analisis sentimen dilakukan untuk melihat dan mengambil informasi berupa opini seseorang dalam bahasa indonesia di Twitter yang ditujukan kepada calon Presiden dan Wakil Presiden Republik Indonesia periode 2019-2024, apakah opini itu masuk kategori opini positif, netral atau negatif. Penelitian ini menggunakan metode *text mining* dan *machine learning* untuk mengambil dan mengklasifikasi polaritas opini dari sumber data. Teknik klasifikasi yang digunakan adalah Naïve Bayes Classifier. Metode ini memiliki kecepatan dan akurasi yang tinggi ketika diaplikasikan dalam basis data yang besar dan data yang beragam [6][7].

Tujuan dari penelitian ini adalah untuk merangkum dan menyimpulkan opini masyarakat yang disuarakan melalui media sosial Twitter terkait pemilihan calon Presiden dan Wakil Presiden Indonesia untuk periode 2019-2024 pasca debat pilpres. Sedangkan manfaatnya adalah membantu masyarakat agar dapat memahami dalam konteks yang lebih luas seperti apa polaritas opini publik terkait calon presiden dan wakil presiden Indonesia tersebut.

2. Metode Penelitian

Langkah-langkah penelitian yang dilakukan digambarkan seperti pada gambar 1.



Gambar 1. Langkah Penelitian.

2.1. Mengumpulkan Data Tweet

Data tweet di ambil melalui Teknik *crawling* di media sosial twitter. Proses *crawling* dilakukan dengan mengembangkan program sederhana berbasis python yang memanfaatkan library Twepy dan Twitter API. Data yang diambil adalah data opini di Twitter yang dituliskan dalam Bahasa Indonesia Mulai tanggal 14 – 17 April 2019. Jumlah data yang diambil untuk masing-masing paslon adalah 100 data opini dengan kata kunci "Jokowi-makruf" dan "Prabowo-sandi". Data diambil secara acak baik yang berasal dari user biasa maupun media online yang menyuarakan opininya di Twitter [8].

Jumlah data opini yang diambil terkait masing-masing paslon digambarkan seperti pada gambar 2.



Gambar 2. Jumlah Data Tweet.

2.2. Preprocessing Data

Preprocessing data dilakukan dalam empat tahapan antara lain melakukan penyelekasiaan opini di database Twitter, hal ini dilakukan agar kalimat opini yang di proses nantinya hanya opini-opini terkait paslon presiden dan wakil presiden republik Indonesia periode 2019-2024. Selanjutnya dilakukan proses cleansing untuk mengurangi noise dari data yang diambil. Selanjutnya dilakukan parsing kalimat. Lalu di tutup dengan proses normalisasi kalimat-kalimat tertentu yang tidak sesuai dengan ejaan yang baku.

2.2.1. Seleksi Data

Pada tahap ini dilakukan seleksi komentar yang mengandung kata atau hashtag(#) JokowiMaaruf, karena Twitter memiliki fungsi retweet, yaitu fitur yang memungkinkan pengguna memberikan komentar terhadap tweet opini seseorang. Komentar tweet dapat mengganggu proses Analisis Sentimen. Oleh karena itu pada tahap preprocessing ini komentar tweet dihapus.

2.2.2. Cleansing Data

Kalimat yang di dapat melalui proses *crawling* umumnya masih terdapat noise yang dapat mengganggu pemrosesan data. Oleh karena itu kita perlu menghilangkan noise tersebut terlebih dahulu. Kata yang dihilangkan adalah karakter HTML, kata kunci, emotikon (ikon emosi), hashtag (#), username (@username), url (http://website.com), dan email (nama@penyedialayanan.com).

2.2.3. Parsing Data

Parsing adalah proses memecah kalimat menjadi sebuah kata dengan melakukan analisa terhadap kumpulan kata dengan memisahkan kata tersebut dan menentukan struktur sintaksis dari tiap kata tersebut [9].

2.2.4. Normalisasi Data

Normalisasi bertujuan untuk menormalkan kalimat sehingga kalimat "gaul" menjadi baku [10], agar dapat dikenali sebagai bahasa yang sesuai dengan KBBI. Yang harus dilakukan dalam proses normalisasi kalimat adalah:

- Memberikan jarak pada tanda baca dan simbol (selain alfabet). Proses ini memastikan setiap kata dan tanda baca sebelum dan sesudahnya dapat di bedakan dengan lebih jelas. Sehingga tanda baca dan simbol selain alfabet tidak menjadi satu dengan kata-kata pada saat proses tokenisasi.
- 2. Mengubah kalimat yang memiliki komposisi huruf yang beragam (huruf besar dan kecil) menjadi seragam (hanya terdiri dari huruf kecil).
- 3. Normalisasi kata dengan aturan seperti yang dituliskan pada Tabel 1.

Tidak Normal	Normal
Akhiran -ny	Akhiran –nya
Akhiran –nk	Akhiran –ng
Akhiran –x	Akhiran –nya
Akhiran –z	Akhiran –s
Akhiran –dh	Akhiran –t
Kata berulang: jalan2	Kata berulang: jalan-jalan
Ejaan: oe	Huruf: u
Eiaan: di	Huruf: i

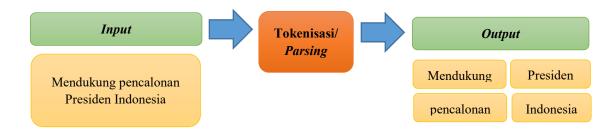
Tabel 1. Aturan Normalisasi Kata [10].

4. Menghilangkan huruf yang berulang. Ketika sedang senang atau kesal, seseorang bebas menuliskan opini berdasarkan emosinya, biasanya seseorang menuliskan dengan mengulang

- huruf yang sama. Contohnya: "kereeen" untuk mengekspresikan kesenangan. Kata berulang seperti "kereeen" akan di normalisasi menjadi "keren".
- 5. Menghilangkan emoticon. Ketika sedang menulis opini (tweet) seseorang kadang salah atau kurang tepat dalam penggunaan emoticon, entah disengaja atau tidak banyak yang melakukannya. Contohnya: Mereka hanya bisa memfitnah karena tidak bisa ketemu fakta buruk :), kata opini fitnah tapi emoticonnya senyum :), dengan begitu emoticon akan mengganggu dalam proses Sentiment Analysis tweet, jadi dalam proses ini emoticon dihapus.

2.3. Tokenisasi

Setelah melakukan normalisasi terhadap data yang dikumpulkan. Langkah selanjutnya adalah melakukan tokenizing atau yang umum disebut parsing. Tahap ini adalah tahap pemotongan string input berdasarkan kata penyusunnya [11]. Pada umumnya setiap kata dalam suatu kalimat dipisahkan menggunakan karakter spasi, sehingga proses tokenisasi ini mengandalkan karakter spasi pada kalimat opini untuk melakukan pemisahan kata dari sekumpulan kata. Contoh proses tokenisasi digambarkan seperti pada gambar 3.



Gambar 3. Proses *Parsing /* Tokenisasi.

2.4. Load Directory

Setelah melakukan tokenisasi terhadap data, langkah selanjutnya adalah melakukan *load dictionary*. Proses *load dictionary* dilakukan agar kita dapat melakukan penggolongan kata sesuai dengan makna asli dari kata yang dimaksud. Pada dasarnya ada beberapa jenis kamus yang dapat digunakan antara lain kamus sentimen positif (kamus kata yang berisi sentiment positif), kamus sentimen negatif, kamus negasi dan sebagainya. Berikut adalah gambaran kamus beserta isinya [10].

- a. Kamus kata kunci positif: jujur, hebat, pandai, baik, keren, cerdas.
- b. Kamus kata kunci negatif: penipu, bohong, korupsi, jahat, jelek.
- c. Kamus negasi: sama, nggak, tidak, bukan, jauh.
- d. Kamus konversi bahasa gaul ke KBBA: sp = siapa, spt = seperti, brp = berapa

2.5. Pembobotan Kata

Setelah kata digolongkan dalam kelompok kata positif, negatif, dan netral menggunakan rancangan kamus yang telah di buat sebelumnya, langkah selanjutnya adalah menentukan bobot nilai yang terkandung di dalam kalimat opini berdasarkan kata penyusunnya. Proses penghitungan dilakukan dengan mengakumulasikan bobot tiap kata yang menyusun suatu kalimat opini. Jika nilai dari kalimat opini tersebut = 1, maka nilai sentiment dari kalimat tersebut adalah positif. Jika nilai

opini = 0, maka nilai sentimen kalimat tersebut adalah netral. Jika nilai opini dalam kalimat tersebut adalah -1, maka nilai sentiment dari kalimat tersebut adalah negatif, seperti yang dituliskan pada tabel 2[10].

Tabel 2. Nilai Sentimen.

Sentimen	Nilai
Positif	1
Netral	0
Negatif	-1

2.6. Klasifikasi

Setelah melakukan pembobotan kalimat opini, langkah selanjutnya adalah melakukan klasifikasi data. Proses klasifikasi dilakukan dengan menggunakan TextBloob 0.15.2. Metode klasifikasi yang digunakan dalam penelitian ini adalah Naïve Bayes Classifier (NBC). NBC adalah metode klasifikasi berdasarkan probabilitas dan Teorema Bayesian dengan asumsi bahwa setiap variabel X bersifat bebas (independence). Dengan kata lain, NBC mengansumsikan bahwa keberadaan sebuah atribut (variabel) tidak ada kaitannya dengan beradaan atribut (variabel) yang lain. Rumus 1berikut merupakan rumus klasifikasi NBC.

$$(A|B) = \frac{p(B|A) x p(A)}{p(B)}$$
 (1)

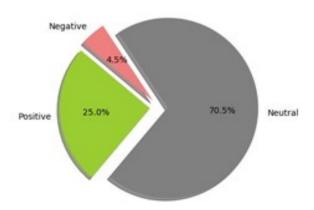
Pada proses klasifikasi data dilakukan dengan menggunakan metode 10-fold *cross validation* [10]. Dataset selanjutnya dibagi menjadi dua, yaitu 10 bagian dengan 9/10 bagian digunakan untuk proses training dan 1/10 bagian digunakan untuk proses testing. Iterasi berlangsung 10 kali dengan variasi data training dan testing menggunakan kombinasi 10 bagian data.

3. Pembahasan

Dataset pada penelitian ini menggunakan format .xml yang dikumpulkan dari Twitter dengan metode Crawling. Data yang diambil hanya tweet opini yang dituliskan dalam bahasa Indonesia, yaitu tweet dengan kata kunci JokowiMakruf untuk PrabowoSandi untuk calon Presiden dan Wakil Presiden Republik Indonesia 2019. Data diambil secara acak tanpa mempertimbangkan pemilik akun baik dari user biasa maupun milik media, baik yang sifatnya *real account* maupun yang terindikasi *fake account*.

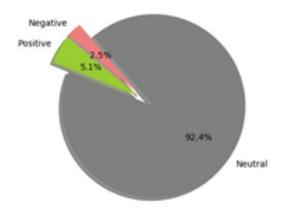
Dataset yang digunakan sebanyak 200 Tweet opini, data dibagi secara seimbang (balanced) setiap kelasnya, karena dengan data yang tidak seimbang (imbalanced), klasifikasi yang dibangun memiliki kecenderungan untuk mengabaikan minority class [10]. Data dibagi menjadi JokowiMakruf 100 Tweet, dan PrabowoSandi 100 Tweet. Pemberian label dilakukan dengan metode Lexicon Based Features.

4. Hasil



Gambar 4. Statistik Polaritas Pasangan Joko Widodo – Ma'aruf Amin.

Grafik pada Gambar 4 mununjukkan hasil polaritas sentimen terhadap capres dan cawapres Joko Widodo dan Ma'aruf Amin. Terlihat sentimen positif terhadap pasangan Joko Widodo – Ma'aruf Amin sebesar 25%. Sentimen negative pasangan ini berada diangka 4.5%, dan sentiment netral berada diangka 70.5%.



Gambar 5. Statistik Polaritas Pasangan Prabowo Subianto – Sandiaga Uno.

Grafik pada Gambar 5 mununjukkan hasil polaritas sentimen terhadap capres dan cawapres Prabowo Subianto dan Sandiaga Uno. Terlihat sentimen positif terhadap pasangan Prabowo Subianto – Sandiaga Uno sebesar 5.1%. Sentimen negative pasangan ini berada diangka 2.5%, dan sentiment netral berada diangka 92.4%.

5. Kesimpulan

Dalam dataset pasangan Joko Widodo dan Ma'aruf Amin, sentimen positifnya paling banyak yaitu 25% dan sentimen negatifnya hanya 4.5%. Sedangkan didataset pasangan Prabowo Subianto dan Sandiaga Uno sentimen positifnya 5.1% dan sentiment negatifnya 2.5%. sehingga dapat disimpulkan, pada media sosial Twitter Jokowi – Ma'aruf lebih banyak di perbincangkan dan di dukung dibandingkan dengan Prabowo - Sandi. Sampai penelitian ini ditulis perkembangan hasil

perhitungan perolehan suara oleh Komisi Pemilihan Umum pun menandakan tren perolehan suara yang positif pada pasangan calon Presiden dan Wakil Presiden Joko Widodo dan Ma'ruf Amin. Terbukti Analisis Sentimen dapat digunakan untuk mengetahui Sentimen masyarakat khususnya user Twitter terhadap calon Presiden dan Wakil Presiden Republik Indonesia Tahun 2019, sehingga diharapkan dapat membantu masyarakat awam untuk mengetahui sentimen masyarakat lainnya terhadap calon Presiden dan Wakil Presiden Republik Indonesia Tahun 2019.

DAFTAR REFERENSI

- [1]. KPU Republik Infonesia, "Penyelenggaraan program penyelenggaraan pemilu tahun 2019." [Daring]. Tersedia pada: https://infopemilu.kpu.go.id/. [Diakses: 08-Feb-2019].
- [2]. Komisioner KPU RI, "Manfaat debat pilpres 2019." [Daring]. Tersedia pada: t https://www.timesindonesia.co.id/read/197488/20190121/220325/masih-belum-tahu-manfaat-debat-pilpres-2019-ini-penjelasan-kpu-ri/. [Diakses: 09-Maret-2019].
- [3]. Freddy. H. Listanto, "Peran Televisi Dalam Masyarakat Citraan Dewasa ini, Sejarah. 6 Perkembangan dan Pengaruhnya", *Jurnal Desain Komunikasi Visual Nirmana*, Vol.1, No.2, (2016)
- [4]. Nurhuda, Faishol, Sari Widya Sihwi, and Afrizal Doewes. "Analisis sentimen masyarakat terhadap calon Presiden Indonesia 2014 berdasarkan opini dari Twitter menggunakan metode Naive Bayes Classifier." ITSmart: Jurnal Teknologi dan Informasi 2.2 (2014): 35-42.
- [5]. Buntoro, Ghulam Asrofi. "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter." INTEGER: Journal of Information Technology 2.1 (2017).
- [6]. Larose, D. T, "Naïve Bayes Estimation and Bayesian Networks, in Data Mining Methods and Models", John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/0471756482.ch5, (2006)
- [7]. McCue, Rita. "A Comparison of the Accuracy of Support Vector Machine and Naive Bayes Algorithms In Spam Classification", (2009)
- [8]. G. A. Buntoro, T. B. Adji, dan A. E. Purnamasari, "Sentiment Analysis Candidates of Indonesian Presiden 2014 with Five Class Attribute," International Journal of Computer Applications, vol. 975, no. 136.2, hlm. 23–29, (2016).
- [9]. A. F. Hadi, D. B. CW, dan M. Hasan, "Text Mining pada Media Sosial Twitter Studi Kasus: Masa Tenang Pilkada DKI 2017 Putaran 2," Seminar Nasional Matematika dan Aplikasinya.
- [10]. G. A. Buntoro, "Analysis Sentimen Calon Gubernur Jawa Timur 2018 di Twitter," Scientico: Computer Science and Informatics journal, vol. 1, no. 2, hlm. 37–45, (2018).
- [11]. Feldman, R & Sanger, J. "The Text Mining Handbook-Advanced Approaches in Analyzing Unstructured Data", USA: New York, (2007)