

Analisis Sentimen Cuitan Twitter Pada Proses Seleksi CPNS Menggunakan Algoritma *Ensembles Vote*

Susi

Teknik Informatika, Universitas Pamulang, Kota Tangerang Selatan
e-mail: susi.skom@gmail.com

Abstrak

Analisis sentimen guna mengetahui komentar pengguna media sosial twitter, dengan menggunakan metode klasifikasi diantaranya 3 (tiga) algoritma lain seperti k-Nearest Neighbor (k-NN), C4.5, Naïve Bayes dan Ensembles Vote. Selanjutnya dilakukan optimasi parameter, misalnya dengan menggunakan cross-validation. Selain itu metode yang digunakan yaitu metode CRISP-DM yang bertujuan untuk memperoleh suatu pola atau pengetahuan dari data yang dimanfaatkan untuk menyelesaikan suatu masalah dan digunakan untuk pengambilan keputusan yang tepat. hasil perhitungan metode k-NN mendapatkan nilai akurasi 61.90% sedangkan hasil perhitungan Naïve Bayes mendapatkan nilai akurasi 100.00%.

Kata Kunci: Twitter, K-NN, C.45, Naïve Bayes, Ensemble Vote.

Abstract

Comment analysis to find out comments from Twitter social media users, using a classification method including 3 (three) other algorithms such as k-Nearest Neighbor (k-NN), C4.5, Naïve Bayes and Ensembles Vote. Furthermore, parameter optimization is carried out, for example by using cross-validation. In addition, the method used is the CRISP-DM method which aims to obtain a pattern or knowledge from the data that is used to solve a problem and is used for making the right decision. the results of the calculation of the k-NN method get an accuracy value of 61.90% while the results of the Naïve Bayes calculation get an accuracy value of 100.00%.

Keywords: Twitter, K-NN, C.45, Naïve Bayes, Ensemble Vote.

I. PENDAHULUAN

Aplikasi media sosial berbasis web dan mobile yang khusus digunakan untuk mengunggah teks saat ini semakin banyak digunakan oleh pengguna mulai dari aktris, seniman, pebisnis, pelaku politik, menyeluruh hingga masyarakat yang gemar maupun hanya menggali informasi hingga instansi-instansi saat ini di Indonesia. Kebutuhan informasi dan kecepatan informasi mengundang para penggiat media sosial tentunya akan meninjau segala informasi, mengolah dalam pikiran dan mengunggah komentar terhadap informasi yang diterima, serta berkomentar sesuai apa yang mereka anggap layak atau tidak layak. Informasi harian yang didapat dan dilihat dari media informasi atau berita yang sedang menjadi bahan perbincangan atau hot issue yang ingin diungkapkan berupa komentar dari informasi didapatkan sebelumnya. *Twitter*, salah satu media social tersebut dapat mereview melalui memberikan komentar, *me-reply/ReTweet* (RT) akun *twitter* lain, *quotes tag*(#), lokasi, ataupun sisipan link kegambar tertentu, menjadi sebuah informasi.

Twitter banyak digunakan akan kemudahannya untuk mengunggah teks dalam media sosial mikro

blog. *Twitter* dalam terbatasnya dalam memberikan komentar dapat dikategorikan sebagai informasi inti suatu *post twit* yang diunggah pada *Twitter*. Komentar yang banyak terhadap *Twitter* oleh akun tertentu, terkenal atau viral karena *follower*-nya memungkinkan RT-nya juga semakin banyak serta semakin informasi tersebut adalah tren, update dan ataupun menjadi sebuah berita yang viral. Untuk itu dalam beberapa penelitian terkait deteksi yang dilakukan oleh beberapa peneliti untuk menggunakan *dataset twitter* banyak digunakan. Penulis dapat mengusulkan suatu teknik tertentu yang digunakan berbasis algoritma, untuk mendeteksi dampak dari suatu hal atau informasi dengan suatu metode yang dapat diteliti dalam penulisan.

Text mining sebenarnya merupakan bagian dari data mining dimana proses yang dilakukan utamanya adalah melakukan ekstraksi pengetahuan dan informasi dari pola-pola yang terdapat dalam sekumpulan dokumen teks menggunakan alat analisis tertentu.

Penerapan metode *machine learning* digunakan untuk mengklasifikasi polaritas suatu berita dari sumber data yang sangat banyak, untuk melakukan

hal itu, bisa menggunakan salah satu fungsi dari *text mining*. Menurut (Anjani, 2015) menjelaskan bahwa *text mining* mempunyai kesamaan dengan *data mining*. Keduanya memiliki tujuan yang sama yaitu untuk memperoleh informasi dan pengetahuan dari sekumpulan data yang sangat besar. Data tersebut bisa berbentuk sebuah *database*. Namun keduanya memiliki perbedaan jenis data

II. METODE PENELITIAN

Analisis Kebutuhan

Metode penelitian yang akan dilakukan terdiri dari beberapa tahapan seperti pengumpulan data, pengolahan data awal, metode yang diusulkan, eksperimen dan hasil pengujian serta evaluasi dan validasi hasil.

Metode Yang Diusulkan

Pada penelitian ini peneliti mengusulkan dalam Analisis komentar guna mengetahui komentar pengguna media sosial *twitter*, dengan menggunakan metode klasifikasi diantaranya 3 (*tiga*) algoritma lain seperti *k-Nearest Neighbor* (k-NN), C4.5, *Naïve Bayes* dan *Ensembles Vote* merupakan metode yang mampu bekerja dengan baik pada data set dengan dimensi tinggi. Selanjutnya dilakukan optimasi parameter, misalnya dengan menggunakan *cross-validation*. Untuk itu penulis melakukan perbandingan terhadap metode-metode tersebut C4.5 dan *k-Nearest Neighbor* (k-NN). Selain itu metode yang diusulkan menggunakan metode CRISP-DM yang bertujuan untuk memperoleh suatu pola atau pengetahuan dari data yang dimanfaatkan untuk menyelesaikan suatu masalah dan digunakan untuk pengambilan keputusan yang tepat dengan langkah-langkah sebagai berikut:

Fase Pemahaman Bisnis (Business Understanding Phase)

Pada tahap ini peneliti membuat satu pertanyaan. Pertanyaannya adalah apa tujuan dari penelitian ini? Jawabannya adalah untuk mengetahui opini masyarakat terkait seleksi CPNS 2021 berdasarkan komentar *twitter* pada media berita *online*, karena banyak pemberitaan yang berbanding terbalik dengan keadaan yang sebenarnya.

Fase Pemahaman Data (Data Understanding Phase)

Pada tahap ini peneliti melakukan pengumpulan data, mengidentifikasi, dan memahami data yang

telah dikumpulkan. Data yang digunakan dalam penelitian ini adalah data yang dikumpulkan dari komentar-komentar artikel berita pada media berita *online*.

Fase Pengolahan Data (Data Preparation Phase)

Langkah pertama yang dilakukan setelah data terkumpul adalah dengan melakukan *cleansing* data dan pengkategorian data komentar kedalam data komentar positif dan data komentar negatif. Lalu ketahap berikutnya *text preprocessing* yaitu mengubah data teks yang tidak terstruktur menjadi data teks terstruktur sehingga data siap digunakan pada teknik pemodelan yang sudah ditentukan. Pada tahap data preparation merupakan sebuah tahap untuk menyiapkan data yang akan diterapkan ke dalam pemodelan, yang sebelumnya berasal dari data mentah awal sampai ke dalam tahap klasifikasi. Dalam tahap ini merupakan tahap yang dilalui setelah *document collection* adalah *preprocessing* bahasa indonesia

Transform Cases

Mengubah kapitalisasi karakter (huruf) menjadi kecil untuk semua kata atau huruf. Setelah melalui tahap transform cases, seluruh isi dokumen menjadi non kapital. Selanjutnya akan diproses pada tahap tokenization.

Tokenization

Proses tokenization dilakukan setelah transform cases. Semua karakter yang tidak diperlukan akan dibuang. Termasuk white space yang berlebihan dan semua tanda baca. Proses ini akan dilakukan terhadap setiap dokumen yang dimasukkan dari *document collection*. Sehingga diperoleh kata yang unik dan dapat merepresentasikan dokumen.

Filter Token (by Length)

Pada tahap ini dilakukan pemilihan token dengan ukuran panjang minimal 3 karakter (huruf), walaupun beberapa diantaranya adalah stopwords. Jika kata yang kurang dari 3 karakter (huruf), namun termasuk dalam stopwords akan tetap dibuang, karena tahap selanjutnya adalah pembuangan stopwords. Misalnya kata seperti "ea", "km", "d", "dr", "e", dan "s" merupakan kata yang kurang dari 3 huruf, namun bukan merupakan stopwords. Hal ini tidak dapat ditangani tahap stopwords. Sedangkan kata seperti "di", "ia", "ke", merupakan kata yang kurang dari 3 huruf namun merupakan stopwords.

Jika kata tersebut terlewat pada tahap ini, masih bisa dibuang pada tahap stopwords.

Stopwords

Tahap stopwords ini akan menyempurnakan tahap filter token by length. Kata yang terdiri lebih dari 3 huruf dan termasuk dalam stopwords akan dibuang. Karena kata tersebut tidak mencerminkan isi dokumen walaupun sering muncul.

Stemming

Semua kata yang telah dipilih untuk menjadi token pada tahapan sebelumnya, akan diubah ke dalam bentuk akar (asal) kata.

Fase Pemodelan (*Modeling Phase*)

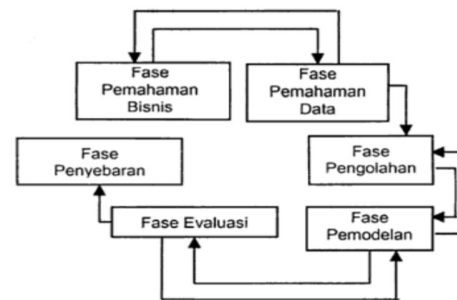
Pada tahap pemodelan ini akan dilakukan teknik pengklasifikasian data yang paling akurat. Untuk membandingkan dan mengkomparasi metode algoritma klasifikasi yang digunakan penulis

Fase Evaluasi (*Evaluation Phase*)

Model yang terbentuk akan diuji menggunakan *confusion matrix* untuk mengetahui tingkat akurasi. *Confusion matrix* akan menggambarkan hasil akurasi mulai dari Analisis positif yang benar, Analisis positif yang salah, Analisis negatif yang benar dan Analisis negatif yang salah. Akurasi akan dihitung dari seluruh hasil Analisis yang benar (baik Analisis positif dan negatif) dibandingkan dengan seluruh data testing. Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan. Pengujian juga diukur dengan menggunakan *ROC Curve*. *ROC Curve* akan menggambarkan kelas positif dalam bentuk kurva. Pengujian dilakukan dengan menghitung nilai *AUC (Area Under Curve)*, semakin tinggi nilai *AUC* dan *ROC Curve*, maka semakin baik pula model klasifikasi yang terbentuk.

Fase Penyebaran (*Deployment Phase*)

Pada *fase deployment* akan dilakukan perancangan aplikasi dengan menggunakan dataset komentar *twitter* ASNBKN.



Gambar 1. Metode Yang Diusulkan Dengan Model CRISP-DM

Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data yang berasal dari *tweet* yang berisi opini dengan kata kunci yaitu “Seleksi CPNS”, melalui akun resmi yaitu @BKNGoid yang dilakukan melalui RapidMiner. Kemudian data tersebut dipisahkan menjadi komentar positif dan komentar negatif. Komentar positif yaitu sentimen dimana pengguna memberikan komentarnya dengan memberikan penilaian secara positif atau memberikan keterangan bahwa setuju dan mendukung seleksi SKD harus diulang. Sedangkan komentar negatif merupakan komentar sebaliknya dari komentar positif yaitu pengguna memberikan komentarnya dengan memberikan ketidaksetujuan terhadap kebijakan para petinggi. Untuk memperoleh data tersebut peneliti mengambil data dari *Twitter* pada bulan November 2021 dengan memilih secara acak dari user pengguna dengan menggunakan *text* bahasa Indonesia.

Penilaian terhadap perilaku pada pengguna *twitter* juga dilakukan untuk menunjang data Analisis seperti *User freq*, *ReTweet*, *To User*, dan *Tema*. pada data-data tersebut dan menjadikan atribut dalam proses dokumen yang akan dilakukan dalam tulisan ini seperti aktif atau pasifnya pengguna *twitter*, *tweet* atau *re-tweet* dan *Tema* positif atau negatif pembahasan yang sedang berlangsung.



Gambar 2. Contoh Komentar Positif Kecurangan Seleksi CPNS 2021



Gambar 3. Contoh Komentar Negatif Kecurangan Seleksi CPNS 2021

Tabel 1. Contoh Data Tweet Positif dan Negatif

No	Data Tweet	Sentimen
1	Ayok dong min mohon tanggapannya,masyarakat juga butuh kejelasan soal isu” yang beredar supaya tidak timbul kegaduhan dan kecurigaan	Positif
2	Min..jangan menjajikan beginilah hidup lu min, kalau digaji pake duit Negara, ada kasus besar pura2 fine aja lu min	Negatif
3	Sudah rahasia umum kalau mau jadi ASN harus setor uang	Negatif

III. HASIL DAN PEMBAHASAN

Pembentukan data sampel tersebut bertujuan untuk mempermudah analisis data dan acuan

pengambilan kebijakan dalam penyelenggaraan Seleksi Calon Pegawai Negeri Sipil (CPNS) 2021. Sebagai penyelenggara Seleksi Calon Pegawai Negeri Sipil (CPNS) 2021, BKN memiliki basis data yang begitu besar. Namun, besarnya data kepesertaan memerlukan perlakuan yang khusus untuk dapat dipergunakan sebagai dasar penyusunan kebijakan. Tes CPNS sempat tertunda karena Covid-19, maka dari itu geliat masyarakat yang ingin ikut tes seleksi CPNS pada tahun 2021 sangat iku andil, bahkan peminatnya sangat besar. Ketika tahapan seleksi berjalan, media bermunculan dengan berbagai pembicaraan tentang adanya kecurangan tersebut, masyarakat yang mendengar dan membaca berita tersebut saling klaim dan saling menjatuhkan dalam komentar dilakukan dimedia sosial seperti Twitter sangat marak beberapa bulan terakhir ini.

Dalam menyikapi hal tersebut, BKN akan usut pada kecurangan tersebut. Menanggapi banyaknya laporan dugaan kecurangan saat seleksi CPNS 2021, Wakil Ketua Komisi II DPR RI Junimart Girsang, mendesak Kemenpan RB dan BKN agar segera melakukan seleksi ulang secara menyeluruh agar masalah kecurangan tersebut bisa teratasi. Pemerintah dan BKN terus berkomitmen menindaklanjuti kecurangan proses seleksi tahap SKD CPNS 2021 dan Badan Kepegawaian Nasional (BKN) menegaskan akan memberi sanksi disiplin berat bagi pegawai yang terbukti terlibat kecurangan.

Sehingga dalam pengumpulan data penulis lakukan sejak periode September 2021 ditarik secara periodik melalui akun official BKN @BKNgoid dan kemudian dibagikan kepada orang lain untuk dilakukan Analisis komentar tersebut positif atau negatif terkait komentar yang di *posting* oleh pengguna *twitter*. Dibagi dalam 200 komentar untuk setiap orang hingga seluruh komentar telah diberikan label atas komentar tersebut.

Berikut merupakan tahapan *text preprocessing* yang digunakan dengan menggunakan *Tools Gata Framework* untuk tahap awal pengolahan data:

