

Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet

1.Dataset

Dataset terdiri dari 400 dokumen tweet dari 4 acara televisi yaitu HitamPutihTransTV, IndonesiaLawyersClubTvOne, KickAndyMetroTV, dan MataNajwaMetroTV. Dataset dilengkapi dengan informasi Jumlah Retweet dari masing-masing dokumen tweet tersebut. Lalu dataset ini tersedia untuk penggunaan pribadi

2. Rumusan Masalah

Dari latar belakang masalah ini dapat dijabarkan sebelumnya oleh penulis diatas, maka dihasilkan rumusan masalah yaitu “Bagaimana penggabungan algoritme K-Nearest Neighbor dan penambahan fitur pembobotan jumlah retweet dalam meningkatkan hasil akurasi dan dapat melakukan klasifikasi dengan hasil yang tepat?”

METODE USULAN

Tahapan proses pada sistem analisis sentimen terhadap tayangan televisi berdasarkan opini masyarakat pada media sosial Twitter menggunakan metode K-Nearest Neighbor dan pembobotan jumlah Retweet diawali dengan proses menghitung pembobotan tekstual, dimulai dari praproses, yang mana dalam proses ini ada beberapa proses yaitu tokenisasi, cleansing, case folding, filterisasi dan stemming. Tahapan selanjutnya yaitu pembobotan kata, dalam proses ini ada beberapa proses yaitu TF, DF, Wtf, IDF, dan TF-IDF.

Tahap selanjutnya dilakukan pembobotan jumlah Retweet (non-tekstual). Pada tahapan ini, jumlah Retweet pada dokumen tweet akan diberikan bobot sesuai dengan jumlah retweet yang ada pada dokumen tersebut dengan proses normalisasi min-max. Normalisasi jumlah retweet dilakukan untuk setiap jumlah retweet pada data uji yang dibandingkan dengan jumlah retweet pada data latih tetangganya. Selanjutnya hasil pembobotan tekstual akan digabungkan dengan

pembobotan jumlah retweet (nontekstual) yang sebelumnya telah dilakukan normalisasi min-max. Hasil dari penggabungan tersebut akan menghasilkan suatu nilai, sehingga dapat diketahui dokumen yang telah dilakukan proses klasifikasi bernilai positif atau negatif.

1. K-Nearest Neighbor

K-Nearest Neighbor (KNN) adalah salah satu metode paling sederhana untuk memecahkan masalah klasifikasi (Adeniyi, Wei, & Yongquan, 2016). Algoritme ini sering digunakan untuk klasifikasi teks dan data (Samuel, Delima, & Rachmat, 2014). Pada metode ini dilakukan klasifikasi terhadap obyek berdasarkan data yang jaraknya paling dekat dengan obyek tersebut (Hardiyanto & Rahutomo, 2016). Klasifikasi teks menggunakan metode KNN akan menghasilkan nilai yang lebih optimal jika menggunakan rumus cosine similarity untuk pembobotan tiap-tiap kata pada dokumen teks yang akan diproses.

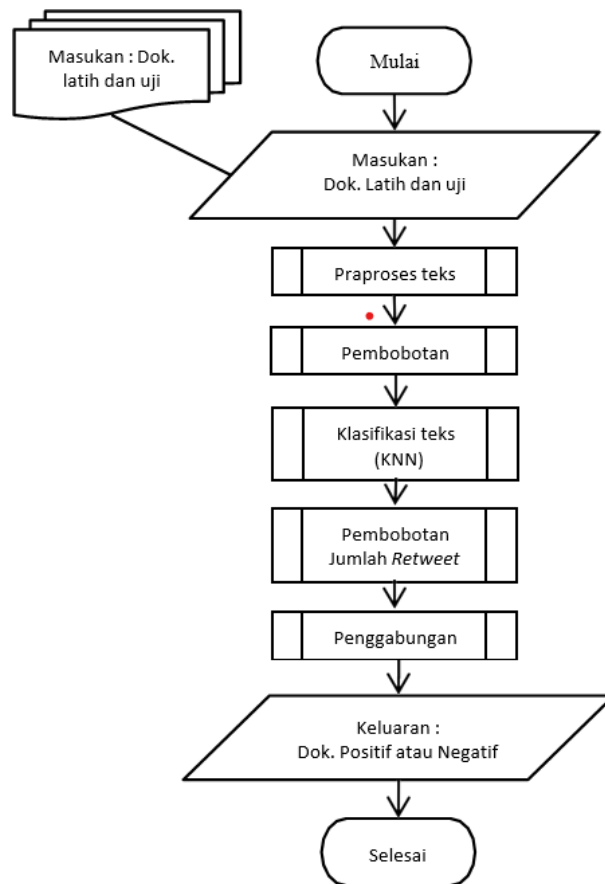
2. Pembobotan Jumlah Retweet (NonTekstual)

Retweet merupakan salah satu fitur yang ada pada media sosial Twitter. Fitur ini berfungsi untuk membagikan atau menyebarkan tweet dari pengguna Twitter lain atau mengacu pada tweet yang didistribusikan kembali. Pembobotan jumlah retweet ini bertujuan untuk menambah nilai positif pada tweet yang memiliki banyak retweet, sehingga dapat mempengaruhi nilai sentimen pada tweet tersebut, dan dapat diketahui dengan jelas tweet apa saja yang memiliki nilai sentimen positif. Jumlah angka retweet dapat memperkuat tweet memiliki sentimen positif (Perdana & Pinandito, 2017). Jumlah retweet merupakan data non-tekstual yang akan dilakukan setelah klasifikasi KNN selesai dilakukan. Bobot yang akan diberikan, berdasarkan jumlah retweet yang terdapat pada dokumen tweet tersebut. Jumlah retweet tersebut akan dilakukan normalisasi menggunakan minmax, agar bobot yang diberikan seimbang dengan bobot teks hasil klasifikasi KNN.

3. Normalisasi Min-max

Normalisasi Min-max merupakan proses transformasi yang mana atribut berupa angka akan diskala ke dalam suatu ukuran yang lebih kecil, seperti antara -1 sampai 1 atau 0 sampai 1. Proses normalisasi tersebut akan memudahkan penelitian, karena data asli akan diganti ke dalam bentuk lain dengan skala yang sama (Maulana, Saepudin, & Rohmawati, 21016). Metode Min-max merupakan metode yang paling sederhana dalam proses transformasi linier terhadap data asli. Setelah dilakukan proses normalisasi Min-max, dapat diperoleh keseimbangan nilai perbandingan antara nilai sebelum dilakukan proses

normalisasi dan nilai setelah dilakukan proses normalisasi.



Gambar 1. Diagram Alur Sistem

VIDEO LINK PENJELASAN

<https://drive.google.com/file/d/1acWv3JsQD0O4D6CFcUDi6jaGprRM3Rak/view?usp=sharing>