

INTISARI

Peringkasan Sentimen Ekstraktif di Twitter Menggunakan *Hybrid TF-IDF* dan *Cosine Similarity*

Oleh

Devid Haryalesmana Wahid
14/371897/PPA/04613

Antusias masyarakat yang memberikan perhatian lebih terhadap akun resmi selebriti di Twitter memunculkan tren penggunaan Twitter sebagai upaya manajemen kesan. Penggalan reaksi masyarakat di media sosial merupakan upaya strategis untuk memperoleh umpan balik, namun tidak mudah dilakukan. Pengguna membutuhkan waktu yang lama untuk membaca ribuan *tweet* sekaligus memilah sentimennya, sehingga dibutuhkan peringkasan sentimen ekstraktif secara otomatis. Penelitian terdahulu umumnya tidak memasukkan informasi sentimen yang terkandung pada sebuah *tweet* sebagai bobot peringkat kalimat, sehingga hasil ringkasan masih berupa topik umum yang dibicarakan masyarakat.

Penelitian ini bertujuan mengkombinasikan metode *SentiStrength*, *Hybrid TF-IDF* dan *Cosine Similarity* untuk mengekstraksi ringkasan sentimen positif dan negatif masyarakat terhadap topik selebriti di Twitter secara otomatis, dengan artis Agnes Monica sebagai studi kasus. Metode *SentiStrength* digunakan untuk mendapatkan skor kekuatan sentimen dan mengklasifikasi *tweet* ke dalam kelas positif, negatif dan netral. *Tweet* bersentimen positif dan negatif diringkaskan dengan cara pemeringkatan *tweet* menggunakan *Hybrid TF-IDF* yang dikombinasi dengan skor kekuatan sentimen, kemudian menghilangkan *tweet* yang mirip menggunakan *Cosine Similarity*.

Hasil pengujian memperlihatkan bahwa kombinasi *SentiStrength*, *Hybrid TF-IDF*, dan *Cosine Similarity* mampu menghasilkan ringkasan sentimen dengan akurasi yang lebih baik dibandingkan menggunakan Hybrid TF-IDF saja, dengan perolehan akurasi rata-rata sebesar 60% dan *f-measure* sebesar 62%. Hal ini disebabkan karena penambahan kekuatan sentimen sebagai bobot peringkasan.

Kata kunci: peringkasan sentimen ekstraktif, analisis sentimen, klasifikasi, peringkasan teks otomatis, *SentiStrength*, *Hybrid TF-IDF*, *Cosine Similarity*.

ABSTRACT

Extractive Sentiment Summarization in Twitter Using Hybrid TF-IDF and Cosine Similarity

By

Devid Haryalesmana Wahid
14/371897/PPA/04613

The using of Twitter by celebrities has become a new trend of impression management strategy. Public reaction mining in social media is a good strategy to obtain feedback, but extracting it is not a trivial matter. User spends a lot of time to reads hundreds of tweets and determine their sentiment polarity. Extractive sentiment summarization machine are needed to address this issue. Previous research generally do not include sentiment information contained in a tweet as weight factor, as a result only general topics of discussion are extracted.

This research aimed to do an extractive sentiment summarization on both positive and negative sentiment mentioning Indonesian celebrity, Agnes Monica, by combining SentiStrength, Hybrid TF-IDF, and Cosine Similarity. SentiStrength is used to obtain sentiment strength score and classify tweet as a positive, negative or neutral. The summarization of positive and negative sentiment can be done by rank tweets using Hybrid TF-IDF summarization and sentiment strength score as additional weight then removing similar tweet by using Cosine Similarity.

The test results showed that the combination of SentiStrength, Hybrid TF-IDF, and Cosine Similarity perform better than using Hybrid TF-IDF only, given an average 60% accuracy and 62% f-measure score. This is due to the addition of sentiment score as a weight factor in sentiment summarization.

Keyword: extractive sentiment summarization, sentiment analysis, classification, text summarization, SentiStrength, Hybrid TF-IDF, Cosine Similarity