

Analisis Sentimen Pada Twitter Terhadap Pergerakan Anti-Vaksin Di Indonesia Menggunakan SentiStrength

Santi Dwi Astuti¹, Ibnu Asror², Moch. Arif Bijaksana³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

⁴Divisi Digital Service PT Telekomunikasi Indonesia

¹santidwiastuti@students.telkomuniversity.ac.id, ²iasror@telkomuniversity.ac.id,

³arifb@telkomuniversity.ac.id,

Abstrak

Internet telah menyediakan teknologi informasi untuk mempelajari masalah-masalah yang berhubungan dengan kesehatan. Munculnya media sosial memberikan cara alternatif untuk bertukar informasi dan membentuk opini tentang isu-isu yang berhubungan dengan kesehatan. Pergerakan anti-vaksin merupakan salah satu penyebab dalam mengurangi tingkat penerimaan vaksin dan dalam meningkatkan wabah yang terjadi dengan opininya. Sentimen masyarakat dapat digunakan sebagai indikator untuk mendapatkan informasi tentang pergerakan anti-vaksin dari media sosial *Twitter*. Tugas Akhir ini dilakukan untuk mengekstrak informasi tentang sentimen di *Twitter*, dilakukan penggalan data kemudian *preprocessing* dengan *tokenizing*, *normalization*, *case folding*, dan *stemming*. Dalam sistem ini menggunakan pendekatan SentiStrength dengan ekstraksi fitur untuk mengkalsifikasi sentimen ke dalam kelas positif, negatif dan netral. Hasil yang diperoleh melalui implementasi dan pengujian sistem menerima data *Twitter* secara keseluruhan berjumlah 558. Penelitian ini memiliki nilai akurasi 70,29% dengan nilai *precision* (positif) 71,72%, *precision* (negatif) 63,64%, *recall* (positif) 89,71%, *recall* (negatif) 33,73%, *f-measure* (positif) 79,71%, dan *f-measure* (negatif) 79,71%. Hasil ini cukup bagus sebagai titik awal untuk penelitian lebih lanjut. Penelitian ini diharapkan dapat memicu penelitian lain dalam analisis sentimen menggunakan SentiStrength, terutama untuk bahasa Indonesia.

Kata kunci : analisis sentimen, anti-vaksin, penambahan teks, SentiStrength, twitter

Abstract

The internet has provided information technology to study health-related problems. The emergence of social media provides an alternative way to exchange information and form opinions about health-related issues. The anti-vaccine movement is one of the causes in reducing the level of acceptance of vaccines and in increasing the outbreaks that occur with his opinion. Community sentiment can be used as an indicator to get information about anti-vaccine movements from *Twitter* social media. This Final Project is done to extract information about sentiments on *Twitter*, extracting data and then preprocessing with tokenizing, normalization, case folding, and stemming. In this system using the SentiStrength approach with feature extraction to calcify sentiments into positive, negative and neutral classes. . The results obtained through system implementation and testing received total *Twitter* data totalling 558 from 407 positive sentiments, 66 negative sentiments, and 65 neutral. This research has an accuracy of 70,23% with precision (positive) 71,72%, precision (negative) 63,64%, recall (positive) 89,71%, recall (negative) 33,73%, f-measure (positive) 79,71%, and f-measure (negative) 79,71%. This result is quite good as a starting point for further research. This research is expected to trigger other research in sentiment analysis using SentiStrength, especially for Indonesian.

Keywords: anti-vaccine, text mining, sentiment analysis, SentiStrength, twitter

1. Pendahuluan

Latar Belakang

Penyakit adalah gangguan kesehatan yang disebabkan oleh bakteri atau virus. Wabah merupakan penyakit menular yang menyebar dengan cepat dan menyerang area luas [14]. Imunisasi secara luas dianggap sebagai salah satu pencapaian terbesar dari kesehatan masyarakat dalam mengurangi kejadian penyakit yang dapat dicegah dengan vaksinasi. Pergerakan anti-vaksin telah ada sejak vaksin dikembangkan pada tahun 1790-an. Pada awal

tahun 2000-an telah mempopulerkan internet sehingga peluang anti-vaksin untuk menyebarkan pesan dan informasi lebih luas untuk menentang vaksinasi di forum berita, situs web dan *platform* media sosial [2]. *Microblog twitter* adalah layanan *platform* media sosial yang banyak digunakan sebagai komunikasi yang cepat dengan keterbatasan maksimum kata yang dikirim menyebabkan pengguna membuat pesan lebih cepat. Pengguna *Twitter* dapat melakukan *posting* status atau pesan yang disebut sebagai *tweet* yang tidak lebih dari 140 karakter.[1]

Tugas akhir ini berfokus pada data *tweet* berbahasa Indonesia, di dalam *tweet* terdapat kosakata non-baku, tidak terstruktur, *emoticon*, *hashtag* dan *link url* sehingga diperlukan teknik *text processing*. Dalam penelitian ada dua pendekatan yang digunakan untuk melakukan analisis sentimen. Yang pertama metode *supervised* adalah klasifikasi dari data *training* yang diberi label secara manual. Metode kedua adalah pendekatan SentiStrength yang salah satu berbasis leksikon dengan memberikan nilai kekuatan sebuah sentimen dan ekstrasi fitur lainnya. Sentimen diklasifikasikan ke dalam polaritas kalimat positif, negatif, dan netral.

Pendekatan SentiStrength dapat melakukan analisis sentimen berbahasa Indonesia yang berbasis leksikon/kamus dengan mendeteksi setiap kata/frasa suatu teks, mengecek *terms* yang mengandung sentimen. SentiStrength ini menggunakan *range* nilai 1 hingga 5 untuk kelas positif, -1 hingga -5 kelas negatif, dan kelas netral.

Topik dan Batasannya

Berdasarkan latar belakang yang sudah dijelaskan, penelitian ini membuat klasifikasi sentimen pada *twitter* terhadap anti-vaksin menjadi tiga kelas yaitu positif, negatif dan netral menggunakan *SentiStrength* pada *tweet* berbahasa Indonesia. Maka rumusan masalah dalam penelitian ini adalah bagaimana hasil penerapan *SentiStrength* terhadap *tweet* berbahasa Indonesia dan hasil evaluasi penggunaan *SentiStrength*. Penelitian ini terdapat beberapa batasan masalah, yaitu *keyword* untuk mengambil *tweet* yang digunakan seputar vaksin dan imunisasi dari tahun 2014-2018 dan *tweet* berbahasa Indonesia. Sebelum proses klasifikasi, dilakukan *preprocessing* yaitu *tokenizing*, *case folding*, *filtering*, dan *stemming*. Ekstrasi fitur yang digunakan yaitu *sentiword*, *emoticon*, *negation*, *question*, dan idiom.

Tujuan

Tujuan dari penelitian tugas akhir ini adalah menganalisa hasil penerapan dan hasil evaluasi *SentiStrength*, mengklasifikasikan kedalam kelas positif, negatif, dan netral, dan membuat visualisasi dalam bentuk grafik untuk mengetahui pergerakan anti-vaksin dari klasifikasi terhadap pelabelan (validasi).

Organisasi Tulisan

Pada tahap pertama dilakukan identifikasi masalah yang menjelaskan mengenai latar belakang, perumusan masalah yang menerapkan *SentiStrength* dan tujuan dari topik yang diambil. Pada tahap kedua dilakukan pencarian materi berisi penelitian sebelumnya yang digunakan sebagai literatur acuan dalam tugas akhir berhubungan dengan *text mining*, *sentiment analysis*, *SentiStrength* dan ilmu semantik (makna kata) untuk bahasa Indonesia. Tahap selanjutnya, dilakukan Perancangan sistem yang telah dibangun sebagai gambaran umum atau disebut alur berjalannya sistem yang berawal dari *input* berupa (*tweet*) hingga *output* berupa hasil kelas sentimen. Kemudian pada tahap Evaluasi menjelaskan hasil dari pengujian yang dilakukan dan analisis terhadap hasil yang dikeluarkan dengan penjelasan skenario pengujian. Tahap terakhir dilakukan menjelaskan hasil penelitian secara keseluruhan dan memberikan saran untuk penelitian selanjutnya.

2. Studi Terkait

2.1 Text Mining

Text mining merupakan bagian dari data mining dan memiliki kesamaan, yaitu *data mining* memiliki *input* data dari data yang sudah terstruktur sedangkan *text mining* data yang tidak terstruktur [13]. *Text mining* atau pemrosesan teks merupakan suatu cara mendapatkan informasi dari suatu dokumen [12]. Dalam penelitian ini menggunakan teknik *text mining* untuk menemukan konten *tweet* terkait vaksin dan imunisasi. Pencarian digunakan dengan kata kunci (*keyword*). Kemudian *tweet* di proses melalui tahapan tertentu yang dapat dilakukan *text mining*, yaitu dilakukan pembakuan kata kemudian *preprocessing* [6][9].

2.2 Sentiment Analysis

Analisis sentimen atau juga disebut dengan *opinion mining* adalah salah satu bagian dari *text mining*. Bidang ini melakukan studi mengenai opini orang-orang, sentimen, evaluasi, tingkah laku dan emosi terhadap suatu entitas seperti produk, layanan, organisasi, individu, permasalahan, topik, acara dan atribut-atributnya. Istilah *sentiment analysis* dimaksudkan sebagai suatu bidang studi untuk menganalisis sentimen pada teks, atau dengan kata lain