

INTISARI

PERBANDINGAN METODE MAXIMUM ENTROPY DAN LEXICON-BASED DALAM ANALISIS SENTIMEN PADA TWEET TENTANG APLIKASI PEDULILINDUNGI

Grace Melira Ruma
17/412638/PA/17957

Saat ini, Twitter adalah salah satu media sosial yang umum digunakan di Indonesia. Sehingga banyak sekali tweets berisi pendapat masyarakat secara umum terhadap suatu produk atau topik dan persoalan di masyarakat. Oleh karena banyaknya jumlah data, maka digunakan *sentiment analysis* untuk memahami dan mengelompokkan informasi dari tweets berbahasa Indonesia dengan *natural language processing*.

Pada penelitian ini, digunakan *Maximum Entropy* (MaxEnt) dan metode *lexicon-based* dalam proses analisis sentimen. Kemudian hasilnya dibandingkan untuk menemukan mana yang paling akurat adalah analisis sentimen berbahasa Indonesia. Kedua metode ini dipilih karena sama-sama cocok untuk data yang berukuran pendek, seperti tweet yang hanya terdiri dari 280 karakter. Namun keduanya memiliki perbedaan yaitu *Lexicon-based* hanya memerlukan kamus *lexicon*, sedangkan MaxEnt membutuhkan training data sebelum bisa digunakan. Training data ini juga sangat memengaruhi hasil dari MaxEnt, dimana semakin besar jumlahnya maka akurasi pun akan semakin baik. Pada penelitian ini dilihat apakah MaxEnt tetap memiliki performa yang baik dengan dataset yang berukuran kecil, dimana training data <1.000 tweets. Untuk metode *lexicon-based* dalam penelitian ini kamus leksikon yang digunakan adalah NRC EmoLex berbahasa Indonesia.

Penilaian yang digunakan adalah akurasi. Akurasi adalah persentase dari hasil analisis yang sentimennya dapat dikenali dengan benar dan tepat. Penilaian akurasi dilakukan dengan *confusion matrix*, yaitu matriks berisi perbandingan hasil dari sistem dan hasil sebenarnya. Pada penelitian ini akurasi yang didapatkan adalah 69.01% untuk MaxEnt dan 42.94% untuk NRC EmoLex.¹

Kata kunci: analisis sentimen, *lexicon-based*, *Maximum Entropy*, NLP

ABSTRACT

COMPARISON BETWEEN MAXIMUM ENTROPY AND LEXICON-BASED METHOD IN SENTIMENT ANALYSIS ON TWEETS ABOUT PEDULILINDUNG

Grace Melira Ruma
17/412638/PA/17957

Twitter is one of the widely used social media in Indonesia. As such there are many tweets containing public opinion on a product or current happenings. Since there are a large amount of data, sentiment analysis is used to understand and classify information from tweets in Indonesian using natural language processing.

In this study, Maximum Entropy (MaxEnt) and lexicon-based method are used in the process of sentiment analysis. Then both methods are compared to find which one have the best accuracy in Indonesian sentiment analysis. These methods were chosen because they are equally suitable for data with short-length, such as tweets as it is consisting of only 280 characters. However, there are also differences in both methods, where Lexicon-based only requires a lexicon dictionary, while MaxEnt requires training data before it can be used. This data training also greatly affects the results of MaxEnt, where the larger the number of the testing data, the better the accuracy. In this study, it will be determined whether MaxEnt can still achieve good performance with a small dataset, in which the training data is <1.000 tweets. For the lexicon-based method in this study, the lexicon dictionary that is used is NRC EmoLex in Indonesian. Accuracy will be used for the assessments. Accuracy is the percentage of the analysis results where sentiments can be identified correctly and precisely. Accuracy assessments will be done using confusion matrix, a matrix of comparison between the result from the system and the real result. In this study, the accuracy for MaxEnt is 69.01% and for NRC EmoLex is 42.94%.

Keywords: sentiment analysis, lexicon-based, Maximum Entropy, NLP