ABSTRACT

SENTIMENT ANALYSIS TO MEASURE POLITICIANS ELECTABILITY USING MULTINOMIAL NAÏVE BAYES

by

Ryan Suryotomo
14/364147/PA/15915

One of politician candidates benchmark to join in election is electability. Recently, the method to measure politicians electability was done conventionally and not objectively, so the result were less representative to the politicians figure. Meanwhile, method that was more modern or objective like sentiment analysis to measure the electability was less used. Twitter and news data were chosen because its could influence politicians public opinions.

Sentiment analysis was performed in this research with tweet and news data for every politicans to measure the electability by using Multinomial Naïve Bayes algorithm. The number of politicians used in this research were 10 politicians that considered as popular politicians in Indonesia. The data set consists of 16.523 training data and 6.550 testing data. The tweet data were collected by using tweetcatcher tool and the news data were collected from 3 news site : tribunnews.com, tempo.co, and viva.co.id by suing scrapper tool in the period of 17th November 2016 until 1st November 2017. Once collected, processing and filtering phase were performed. Then, top-n word features were performed by using chi square and TF-IDF algorithm. The next phase was forming classification models and testing process that compared electabilities result of each politicians with chi square and TF-IDF feature selection or without feature selection.

The result of this research showed that average performance of chi square features selection model was higher with 85,24% accuracy, 88,84% precision, 91,65% recall and 90,17% f-measure compared to TF-IDF features selection which had average value of 78,11% accuracy , 87,41% precision , 87,79% recall and 87,54 f-measure and without features selection model which had average value of 74,69% accuracy , 87,40% precision , 84,88% recall and 84,72% f-measure.

Keyword : sentiment analysis, electability, multinomial naïve bayes, chi square, TF-IDF

INTISARI

ANALISIS SENTIMEN UNTUK MENGETAHUI ELEKTABILITAS TOKOH POLITIK MENGGUNAKAN METODE *MULTINOMIAL NAÏVE BAYES*

Oleh

Ryan Suryotomo
14/364147/PA/15915

Salah satu tolak ukur kandidat tokoh politik yang akan mengikuti pilgub, pilkada atau pemilu adalah elektabilitas. Sekarang ini, metode untuk mengukur elektabilitas tokoh politik masih dilakukan secara konvensional dan tidak objektif sehingga hasilnya kurang merepresentasikan tokoh politik tersebut.Sementara metode yang lebih modern dan objektif seperti analisis sentimen menggunakan data Twitter dan berita untuk mengukur elektabilitas masih sedikit dilakukan. Data Twitter dan berita dipilih karena dapat mempengaruhi opini publik tokoh politik.

Pada penelitian ini dilakukan analisis sentimen menggunakan data tweet dan berita dari masing-masing tokoh politik untuk mengetahui elektabilitasnya menggunakan metode *Multinomial Naïve Bayes*. Tokoh politik yang digunakan dalam penelitian adalah 10 tokoh politik yang dianggap populer di Indonesia. Dataset yang digunakan berjumlah 16.523 data *training* dan 6.550 data *testing*.Data tweet didapatkan menggunakan *tool tweetcatcher* dan berita didapatkan dari 3 situs berita di Indonesia yaitu *tribunnews.com*, *tempo.co*, dan *viva.co.id* menggunakan *tools scrapper* dalam kurun waktu 17 November 2016 sampai 1 November 2017. Setelah data terkumpul, dilakukan tahap *preprocessing* dan *filtering*. Lalu dilakukan seleksi *top-n* kata fitur menggunakan metode *chi square* dan TF-IDF. Selanjutnya adalah pembentukan model klasifikasi dan proses testing dengan membandingkan hasil elektabilitas tiap tokoh politik tanpa seleksi fitur dan dengan seleksi fitur *chi square* dan TF-IDF.

Hasil penelitian ini menunjukan bahwa nilai performa model menggunakan metode seleksi fitur *chi square* lebih tinggi dengan rata-rata nilai akurasi 85,24% , presisi 88,84% , *recall* 91,65% dan *f-measure* 90,17% dibandingkan dengan menggunakan metode seleksi fitur TF-IDF dengan rata-rata nilai akurasi 78,11% , presisi 87,41%, recall 87,79% dan *f-measure* 87,54% serta jika dibandingkan tanpa seleksi fitur dengan nilai rata-rata akurasi 74,69% , presisi 87,40%, *recall* 84,88% dan *f-measure* 84,72%.

Kata kunci: analisis sentimen, elektabilitas, multinomial naïve bayes, *chi square*, TF-IDF