

INTISARI

Deteksi Sarkasme Pada Komentar Twitter Berbahasa Indonesia Menggunakan Metode *Long Short-Term Memory*

Oleh:

Suko Tyas Purnanda
17/418668/PPA/05452

Twitter merupakan media sosial yang memberikan akses kepada penggunanya untuk menanggapi kejadian yang sedang diperbincangkan oleh publik. Seringkali pengguna Twitter menyampaikan pesan secara implisit dengan sarkasme, sehingga menyebabkan sebagian besar algoritma analisis sentimen tidak tepat dalam memberikan label.

Penelitian ini bertujuan untuk mendeteksi sarkasme pada komentar Twitter berbahasa Indonesia dengan metode *Long Short-Term Memory* (LSTM). Data yang digunakan dikumpulkan dengan metode *scraping*, kemudian diberikan nilai vektor menggunakan metode *Word2Vec*, dan diklasifikasikan dengan metode LSTM. Pada tahap pengujian, metode LSTM dibandingkan dengan metode *Naïve Bayes* dengan ekstraksi fitur TF-IDF untuk dicari kombinasi nilai akurasi, *precision*, *recall*, dan *f1-score* yang terbaik.

Pengujian menunjukkan bahwa metode LSTM menghasilkan nilai yang lebih baik daripada *Naïve Bayes*. Bidirectional LSTM sebagai salah satu variasi metode LSTM memiliki akurasi tertinggi 82.13%, *precision* 69.06%, *recall* 58.14%, dan *f1-score* 61.31%. Metode *Naïve Bayes* dengan nilai $\alpha = 0.0001$ memiliki akurasi tertinggi 80.16%, *precision* 67.03%, *recall* 46.92%, dan *f1-score* 55.20%.

Kata kunci: sarkasme, twitter, lstm, *naïve bayes*.

ABSTRACT

Sarcasm Detection on Indonesian Language Tweets from Twitter Using Long Short-Term Memory Method

Oleh:

Suko Tyas Parnanda
17/418668/PPA/05452

Twitter is among the top social media today which provides its users access to express or response towards something that is being discussed in public. Oftentimes, Twitter's users post their tweets in an implicit way by using sarcasm, making it an issue for most sentiment analysis algorithm because they often misclassify a tweet due to the sarcasm content.

This study aims to address the problem by detecting sarcasm in an Indonesian tweet by using Long Short-Term Memory (LSTM). The data used in this study is collected using scraping method, then will be assigned a vector for each word using Word2Vec method, and lastly it will be classified using LSTM method. In the evaluation step, LSTM is compared to Naïve Bayes with feature extraction of TF-IDF to find the best combination value for their accuracy, precision, recall, and f1-score.

Evaluation step shows that LSTM produces better results than Naïve Bayes. Bidirectional LSTM, one of LSTM method variation, has highest accuracy of 82.13%, precision of 69.06%, recall of 58.14%, and f1-score of 61.31%. On the other hand, Naïve Bayes with $\alpha = 0.0001$ has highest accuracy of 80.16%, precision of 67.03%, recall of 46.92%, and f1-score of 55.20%.

Keywords: sarcasm, twitter, lstm, naïve bayes.

