

INTISARI

ANALISIS SENTIMEN PERUSAHAAN GOOGLE MENGGUNAKAN *BI-DIRECTIONAL LONG SHORT TERM MEMORY* PADA DATA TIDAK SEIMBANG

Oleh

Visco Ooi'ncu
(17/414661/PA/18161)

Perkembangan teknologi telah mengubah cara masyarakat mengonsumsi informasi. Media informasi yang dulunya berupa media cetak seperti koran dan majalah telah berubah menjadi digital. Selain itu, informasi yang diterbitkan oleh ratusan penerbit dapat diakses dalam waktu yang singkat. Melakukan analisis manual pada data tekstual yang terbit setiap saat seperti itu merupakan pekerjaan yang sulit dan membutuhkan waktu yang sangat banyak. Analisis sentimen merupakan proses analisis yang diotomatisasi dengan menggunakan komputasi untuk menemukan sentimen positif, netral atau negatif dari data tekstual. Analisis sentimen digunakan secara luas untuk mendapatkan persepsi atau opini dari komentar sosial media, hasil survei kuesioner dan ulasan suatu produk.

Dalam beberapa tahun ini, *deep learning* mendapat perhatian para akademisi karena performa modelnya yang tinggi dalam menangani berbagai hal. Salah satu model Deep Learning yang cukup populer digunakan dalam menangani data teks seperti analisis sentimen adalah *Long Short Term Memory*, selanjutnya disingkat LSTM. LSTM merupakan perkembangan dari *Recurrent Neural Network* yang mengatasi masalah kesulitan dalam menyimpan informasi yang terlalu lama (*long-term dependency problem*). LSTM ini dapat digunakan untuk melakukan analisis pada informasi teks seperti judul berita dengan performa yang tinggi karena memiliki *gates* yang dapat menyimpan dan menyeleksi informasi. Pada penelitian ini, dilakukan pembentukan model klasifikasi sentimen dari judul berita Google menggunakan LSTM dan variannya yaitu Bi-LSTM, *Stacked LSTM* dan *Stacked Bi-LSTM*. Selain itu, teknik *oversampling* digunakan dalam menangani data yang tidak seimbang. Berdasarkan hasil pembentukan model menggunakan beberapa kombinasi dimensi *word embedding*, jumlah neuron dan penggunaan *dropout*, diperoleh hasil bahwa model pada *stacked Bi-LSTM* pada data seimbang memiliki performa paling baik dengan nilai *f1-score* (micro) sebesar 89,66% ; *f1-score* (macro) sebesar 87,26% dan *f1-score* (weighted) sebesar 89,57%.

Kata kunci : analisis sentimen, *deep learning*, *long short-term memory*, *bidirectional LSTM*, data tidak seimbang

ABSTRACT

GOOGLE COMPANY SENTIMENT ANALYSIS USING BI-DIRECTIONAL LONG SHORT TERM MEMORY FOR IMBALANCED DATA

By

Visco Ooi'nco
(17/414661/PA/18161)

Technological developments are changing the way we consume information. Informative media that used to be in the form of print media such as newspapers and magazines have turned into digital. In addition, we can also access information published by hundreds of publishers in a short time. Performing manual analysis on such huge amount of textual data is a difficult task and takes a lot of time. Sentiment analysis is an automated analysis process using computation to find positive, neutral or negative sentiments from textual data. Sentiment analysis is widely used to get perceptions or opinions from social media comments, questionnaire survey results and product reviews.

In recent years, deep learning has received the attention of academics because of high performance of its models in dealing various things. One of deep learning models that is quite popular used in handling text data such as sentiment analysis is Long Short Term Memory, namely LSTM. LSTM is a type of Recurrent Neural Network architecture that is refined to overcome the problem of difficulty in storing information that is too old (long-term dependency problem). LSTM can be used to analyze textual information such as news headlines with high performance because it has gates that store and filter information. In this study, sentiment classification models were developed from Google news headline using LSTM and its variants, namely Bi-LSTM, Stacked LSTM and Stacked Bi-LSTM. In addition, oversampling technique is used in handling imbalanced data. After developing models using several combinations of word embedding dimensions, number of neurons and the use of dropout, the results show that stacked Bi-LSTM model on balanced data has the best performance with the f1-score (micro) value of 89.66%; f1-score (macro) of 87.26% and f1-score (weighted) of 89.57%.

Keywords : sentiment analysis, deep learning, long short-term memory, bidirectional LSTM, imbalanced data