

ABSTRAK

Hans Reinaldy (01032170002)

ANALISIS SENTIMEN PADA PENGGUNA TWITTER TERHADAP PEMILU AMERIKA SERIKAT 2020 MENGGUNAKAN *DEEP LEARNING* DENGAN ALGORITMA *LONG SHORT-TERM MEMORY*

Skripsi, Fakultas Sains dan Teknologi (2021).

(xiii + 55 halaman; 5 tabel; 23 gambar; 4 lampiran)

Penggunaan platform *microblogging* Twitter tidak hanya untuk mengekspresikan keseharian seseorang namun juga pendapat mereka terhadap suatu barang, tokoh, ataupun peristiwa. Hal inilah yang membuat Twitter dianggap oleh ilmuwan dan ahli statistik sebagai sumber yang relevan untuk merepresentasikan opini publik terhadap sesuatu atau seseorang pada dunia nyata. Pemilu Amerika Serikat 2020 merupakan peristiwa yang penting khususnya bagi warga Amerika Serikat, sehingga ada banyak *tweets* mengenai pemilu tersebut. *Tweets* menjadi sumber informasi atau data yang dapat digunakan untuk merepresentasikan sentimen orang-orang terhadap pemilu tersebut. Analisis Sentimen untuk mengklasifikasikan *tweets* dilakukan dengan metode *supervised Deep Learning* dengan algoritma *Long Short-Term Memory*. Terdapat tiga model yang dikembangkan dengan arsitektur yang mirip, dengan perbedaan utama ialah: jumlah *vocabs*, jumlah iterasi, dan besar *training dataset*. Ketiga model menggunakan *training dataset* yang sama, dengan ukuran data yang berbeda. Model 1 dilatih dengan 526.000 data, Model 2 dilatih dengan 126.000 data, dan Model 3 dilatih dengan 1.578.600 data. Model 1 dilatih sebanyak 8 iterasi dan mendapatkan 79,56% *accuracy*, 0,4342 *loss*, dan bersifat *good-fit*. Model 2 dilatih sebanyak 50 iterasi dan mendapatkan 92,45% *accuracy*, 0,1773 *loss*, tetapi sangat *overfit*. Model 3 dilatih sebanyak 8 iterasi dan mendapatkan 85,41% *accuracy*, 0,3342 *loss*, dan sedikit *overfit*. Ketika dibandingkan dengan survei *Popular Vote* oleh *Cook Political*, baik Model 1 dan Model 3 menunjukkan hasil prediksi yang mirip. Hasil Survei menunjukkan bahwa Joe Biden memenangkan 51,3% suara dan Donald Trump memenangkan 46,9% suara. Model 1 memprediksi bahwa Joe Biden memiliki 52,17% sentimen positif sedangkan Donald Trump memiliki 47,83% sentimen positif. Model 3 memprediksikan Joe Biden memiliki 52,55% sentimen positif dan Donald Trump memiliki 47,56% sentimen positif. Tetapi, Model 2 menunjukkan hasil yang sebaliknya dengan Donald Trump mendapat 58,22% sentimen positif dan Joe Biden 41,26% sentimen positif, yang dikarenakan model *overfitting*. Lebih lanjut, pada visualisasi *wordcloud*, kata “win” adalah *most frequent word* pada *wordcloud* Joe Biden, hal ini mengindikasikan adanya harapan untuk suatu perubahan politik.

Kata Kunci : Analisis Sentimen, *Deep Learning*, *Long Short-Term Memory*, *Natural Language Processing*

Referensi : 28 (2009-2020)

ABSTRACT

Hans Reinaldy (01032170002)

TWITTER SENTIMENT ANALYSIS ON U.S. 2020 PRESIDENT ELECTION USING DEEP LEARNING WITH LONG SHORT-TERM MEMORY ALGORITHM

Thesis, Faculty of Science and Technology (2021).

(xiii + 55 pages; 5 tables; 23 figures; 4 appendices)

As a microblogging platform, Twitter has been widely used by people not only to express their daily activities but also their opinion towards an object, a public figure, or an event. Because of that, scientist and statistician stated that Twitter is a valid source to represent public opinion towards something or someone. The 2020 United States Presidential Election is one of the momentous events that many have waited for so long, especially for the United States citizen and as one of the momentous events, undoubtedly the number of tweets regarding the issue is great. Those tweets are important source of information or data that can be used to represent people sentiment about the election. The Sentiment Analysis classifies tweets according to their polarity (positive or negative) in a supervised way using Deep Learning method with Long Short-Term Memory algorithm. There are three models built upon the similar architecture, with the main differences are the number of vocabs — which affects the embedding layer dimension, number of iterations, and size of training datasets. Three models are trained using the same training dataset with different data sizes. Model 1 trained using 526,000 data, Model 2 trained using 126,000 data, and Model 3 trained using 1,578,600 data. Model 1 trained for 8 iteration achieved 79.56% accuracy, 0.4342 loss, and good-fitted. Model 2 trained for 50 iteration achieved 92.45% accuracy, 0.1773 loss, but highly overfitted. Model 3 trained for 8 iteration and achieved 85.41% accuracy, 0.3342 loss, and slightly overfitted. When compared with Cook Political Popular Vote survey, both Model 1 and Model 3 show similar sentiment predictions. The survey claims Joe Biden wins 51.3% votes while Donald Trump wins 46.9% votes, Model 1 predicted Joe Biden has 52.17% positive sentiments and Donald Trump has 47.83% positive sentiments, and Model 3 predicted Joe Biden has 52.55% positive sentiments and Donald Trump has 47.56% positive sentiments. However, Model 2 shows that Donald Trump has 58.22% positive sentiments while Joe Biden has 41.26% positive sentiments due to overfitted model. Furthermore, the *wordcloud* visualization found that, Twitter’s users hope for a political change, indicated by “win” word appears as the most frequent word in Joe Biden *wordcloud*.

Keywords : Deep Learning, Long Short-Term Memory, Natural Language Processing, Sentiment Analysis

Reference : 28 (2009-2020)