

## ABSTRAK

Dennis Christian (01032180010)

**PENGUNAAN *MACHINE LEARNING* UNTUK PREDIKSI HASIL PEMILU DAN MEMBANDINGKAN MEREK MENGGUNAKAN ANALISIS SENTIMEN DATA TWITTER**  
Skripsi, Fakultas Sains dan Teknologi (2022).

(xiv + 47 halaman; 35 gambar; 10 tabel; 4 lampiran)

Kemudahan pengambilan data di era digitalisasi ini memungkinkan *machine learning* bisa dilatih dengan *dataset* yang sangat besar sehingga teknologi *machine learning* semakin maju. Sebagai contoh, data Twitter bisa digunakan untuk analisis sentimen, yaitu mengklasifikasi tulisan ke dalam kategori positif, negatif, dan bisa juga netral. Model *machine learning* yang diuji adalah *Naive Bayes* dan dua jenis *deep learning*, yaitu *Convolutional Neural Network (CNN)*, dan *Recurrent Neural Network (RNN)*. *Deep learning* merupakan bagian dari *machine learning* yang menggunakan algoritma *neural network*. Ketiga model tersebut bekerja dengan cara yang berbeda, *Naive Bayes* memprediksi kategori suatu kalimat dengan menghitung probabilitas menggunakan rumus teorema Bayes, CNN melakukan konvolusi data *input* terlebih dahulu sebelum dimasukkan ke *neural network*, sedangkan RNN memiliki memori yang memungkinkan untuk mengingat *input* sebelumnya untuk diproses bersama *input* setelahnya. Untuk membandingkan performa ketiga model, *training* dilakukan menggunakan *dataset* Bahasa Inggris dengan kategori positif dan negatif yang di-*download* dari thinknook. Dari ketiga model tersebut, penulis memilih RNN karena memiliki akurasi paling tinggi. Untuk meningkatkan lagi akurasi, penulis menyortir kata-kata dalam *dataset training* dan *testing* menggunakan Stanford NER untuk menghapus kata-kata dengan topik yang tidak terlalu mengandung sentimen. Penulis juga melakukan penggeseran *threshold output*. Setelah prosedur-prosedur optimasi tersebut dilakukan, didapatkan akurasi 78,11% saat model diuji menggunakan data Twitter yang sudah dilabeli oleh penulis secara manual. Model yang akurasi sudah meningkat ini kemudian digunakan untuk menganalisis sentimen Twitter terhadap merek Coca-Cola, Pepsi, Steam, Epic Games, dan Nokia namun hasil menunjukkan bahwa persentase sentimen positif terhadap suatu merek tidak terlalu berkorelasi dengan kinerja merek tersebut. Kemudian model digunakan untuk menganalisis sentimen Twitter terhadap calon presiden Amerika Serikat dan Indonesia. Didapatkan bahwa analisis sentimen ini dapat memprediksi calon mana yang dipilih paling banyak.

Kata Kunci : Analisis Sentimen, *Machine Learning*, RNN, Twitter

Referensi : 22 (2011–2021)

## ABSTRACT

Dennis Christian (01032180010)

### **IMPLEMENTATION OF MACHINE LEARNING FOR ELECTION RESULTS PREDICTION AND COMPARING BRANDS USING SENTIMENT ANALYSIS OF TWITTER DATA**

Thesis, Faculty of Science and Technology (2022).

(xiv + 47 pages; 35 figures; 10 tables; 4 appendices)

The ease of data retrieval in this digitalization era allows machine learning to be trained with very large datasets so that machine learning technology is advancing. For example, Twitter data can be used for sentiment analysis, which is to classify posts into positive, negative, and neutral categories. The machine learning model tested is Naive Bayes and two types of deep learning, namely Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). Deep learning is part of machine learning that uses neural network algorithms. The three models work in different ways, Naive Bayes predicts the category of a sentence by calculating the probability with Bayes theorem formula, CNN convolutes the input data first before it is entered into the neural network, while RNN has a memory that allows it to remember the previous input to be processed with the next input. To compare the performance of the three models, training was conducted using an English dataset with positive and negative categories that downloaded from thinknook. Of the three models, the author chose RNN because it has the highest accuracy. To further improve its accuracy, the author sort the words in the training and testing dataset using Stanford NER to remove words with less sentimental topics. The author also shifts the output threshold. After the optimization procedures were carried out, an accuracy of 78.11% was obtained when the model was tested using Twitter data that had been manually labeled by the author. The model with increased accuracy is then used to analyze Twitter sentiment towards the Coca-Cola, Pepsi, Steam, Epic Games, and Nokia brands, but the results show that the percentage of positive sentiment towards a brand is not highly correlated with the brand's performance. Then the model is used to analyze Twitter sentiment towards the presidential candidates of the United States and Indonesia. It was found that this sentiment analysis can predict which candidate is chosen the most.

Keywords : Sentiment Analysis, Machine Learning, RNN, Twitter

Reference : 22 (2011–2021)