



INTISARI

Saat ini teknologi *machine learning* telah banyak digunakan dalam berbagai bidang mulai dari mesin pencarian (*search engine*) seperti Google Search, analisis data untuk merekomendasikan suatu produk, konten iklan yang akan ditampilkan, asisten virtual (*virtual assistant*) seperti "OK Google", pembuatan *otonomous vehicle*, hingga analisis suatu media sosial. Telah banyak metode *machine learning* yang dikembangkan, yakni metode *Supervised Machine Learning*, *Unsupervised Machine Learning*, serta *Semi-Supervised Machine Learning*. Algoritme yang ada atau disediakan untuk *machine learning* pun sangat banyak, salah satu yang penulis gunakan adalah algoritme *XGBoost Supervised Machine Learning*.

Pada penelitian ini, dilakukan sebuah perancangan suatu sistem *classifier* agar dapat melakukan prediksi akun bot pada media sosial Twitter dengan kata kunci yang digunakan, dalam penelitian ini yaitu "#pilpres". *Classifier* yang dibuat berbasis *machine learning* dengan menggunakan algoritme XGBoost.

Bot detector akan bertugas untuk melakukan deteksi jenis akun pada media sosial Twitter. Kemudian akan diukur performa dari *classifier* dan dilakukan perbandingan antara performa satu *classifier* dalam menganalisis 3000 twit dan performa tiga *classifier* dengan jumlah twit yang sama.

Sistem terdistribusi ini bertugas untuk melakukan deteksi jenis akun pada media sosial Twitter. Kemudian performa dari *classifier* diukur dan kemudian dilakukan perbandingan antara performa sistem yang menggunakan satu *classifier* dalam menganalisis 3000 twit dan performa sistem yang menggunakan tiga *classifier* secara bersamaan dengan jumlah twit yang sama. Hasil dari pengujian menunjukkan bahwa sistem dengan satu *classifier* memerlukan waktu 55,44 detik dan mengkonsumsi RAM sebanyak 6,26 MB. Sedangkan bila menggunakan tiga *classifier*, sistem menghabiskan waktu rata-rata 18,53 detik dan mengkonsumsi RAM sekitar 5.6 MB.

Kata kunci: *Machine Learning, XGBoost, Text Processing, Supervised Learning, Twitter, Stemming, Vectorizing*



ABSTRACT

Nowadays machine learning technology has been widely used in various fields ranging from search engines such as Google Search, data analysis to recommend a product, ad content to be displayed, virtual assistants such as "OK Google", autonomous creation vehicle, to the analysis of a social media. Many machine learning methods have been developed, namely the Supervised Machine Learning method, Unsupervised Machine Learning, and Semi-Supervised Machine Learning. There are many algorithms available for machine learning, one of which I use is the XGBoost Supervised Machine Learning algorithm.

In this study, a classifier system was designed in order to predict bot accounts on Twitter social media with the keywords used, in this study "#pilpres". Classification was made based on machine learning using XGBoost algorithm.

The performance of the classifiers and the comparison between the performance of one classifier in analyzing 3000 tweets and the performance of three classifiers with the same number of tweets were measured. The results of the test showed that the system with one classifier took 55.44 seconds and consumed 6.26 MB of RAM. While the system using three classifier spent an average of 18.53 seconds and consumed about 5.6 MB of RAM.

Keywords: Machine Learning, XGBoost, Text Processing, Supervised Learning, Twitter, Stemming, Vectorizing