

ABSTRAK

ST. NUR HALISAH DULI. Analisis Machine Learning dan Deep Learning Terhadap Ukuran Dataset Study Kasus Analisis Sentimen Text (dibimbing oleh Fahrin Irhamna Rahman S.Kom.,M.T. dan Titin Wahyuni S.Pd.,M.T).

Penelitian ini bertujuan untuk menentukan ukuran dataset yang optimal dalam penggunaan machine learning dan deep learning. Metode yang digunakan dalam machine learning adalah Naive Bayes, Regresi Logistik, dan Support Vector Machine. Sedangkan untuk Deep Learning, metode yang digunakan adalah Convolutional Neural Network. Data yang digunakan dalam penelitian ini adalah data ulasan yang diperoleh dari Google Maps untuk beberapa tempat wisata, yaitu Bugis Waterpark, Akkarena, Tanjung Bayang, Pantai Bosowa, dan Wisata Kebun. Penelitian ini melibatkan tahap preprocessing untuk membersihkan data, seperti cleaning, case folding, stopwords, tokenizing dan stemming. Pengujian data dilakukan menggunakan 9 set data dengan ukuran yang berbeda, yaitu 4500, 4000, 3500, 3000, 2500, 2000, 1500, dan 500. Pembagian data dilakukan antara data latih dan data uji dengan perbandingan 90:10, 80:20, dan 70:30. Pengujian akurasi dengan menggunakan empat metode dan sembilan ukuran data menunjukkan variasi dalam tingkat akurasi. Hasil penelitian menunjukkan bahwa metode regresi logistik dengan ukuran data 1000 dan pembagian data 90:10 memberikan akurasi tertinggi, yaitu 85%.

Kata Kunci : Analisis sentimen, Naïve Bayes, Regresi Logistik, Support Vector Machine, Convolutional Neural Network.

ABSTRACT

ST. NUR HALISAH DULI. *Analysis of Machine Learning and Deep Learning on Dataset Size: A Case Study of Sentiment Text Analysis (supervised by Fahrira Irhamna Rahman S.Kom.,M.T. and Titin Wahyuni S.Pd.,M.T).*

This research aims to determine the optimal dataset size in the usage of machine learning and deep learning. The methods employed in machine learning include Naive Bayes, Logistic Regression, and Support Vector Machine, while for Deep Learning, Convolutional Neural Network is utilized. The data used in this study consists of reviews obtained from Google Maps for several tourist attractions, namely Bugis Waterpark, Akkarena, Tanjung Bayang, Pantai Bosowa, and Wisata Kebun. This research involves preprocessing steps to clean the data, such as cleaning, case folding, stopwords removal, tokenizing, and stemming. Data testing is conducted using 9 sets of data with different sizes, namely 4500, 4000, 3500, 3000, 2500, 2000, 1500, and 500. The data is divided between training and testing data with ratios of 90:10, 80:20, and 70:30. Accuracy testing using four methods and nine data sizes indicates variation in accuracy levels. The research results indicate that the logistic regression method with a dataset size of 1000 and a 90:10 data division ratio yields the highest accuracy, reaching 85%.

Keywords: *Sentiment analysis, Naïve Bayes, Logistic Regression, Support Vector Machine, Convolutional Neural Network.*