

**DAMPAK UKURAN DATASET PELATIHAN TERHADAP
PERFORMA *LONG SHORT TERM MEMORY*
(*LSTM*) *NETWORK* DALAM KONTEKS
HARGA SAHAM**

ABSTRAK

Performa jaringan *LSTM* umumnya tidak mudah diketahui sebelum proses pelatihan dan pengujian jaringan tersebut sudah dilakukan. Salah satu faktor yang mempengaruhi proses tersebut adalah ukuran dataset pelatihan yang digunakan oleh jaringan *LSTM*. Untuk melihat dampak faktor tersebut, pengujian dilakukan dengan beberapa variasi ukuran dataset awal yang digunakan untuk pelatihan dan yang digunakan untuk pengujian. Kasus yang digunakan dalam penelitian ini adalah kasus prediksi harga saham, dimana harga saham masa lalu digunakan untuk mencari harga saham masa depan. Empat bank dipilih sebagai sumber dataset awal, dan setiap dataset awal terdiri dari enam fitur dataset, yaitu: *Open*, *High*, *Low*, *Close*, *Adj. Close*, dan *Volume*. Lima variasi ukuran rasio dataset digunakan, yaitu: 70/30, 75/25, 80/20, 85/15, dan 90/10. Hasil yang didapatkan dari pengujian menunjukkan rasio dataset 90/10 memberikan performa *LSTM Network* yang paling bagus untuk hampir semua fitur dataset yang diuji.

Kata kunci: Ukuran dataset pelatihan, Jaringan *LSTM*, Harga saham, Bank.

**IMPACT OF TRAINING DATASET SIZE ON LONG SHORT TERM
MEMORY (LSTM) NETWORK PERFORMANCE
IN THE CONTEXT OF STOCK PRICE**

ABSTRACT

It is difficult to tell how an LSTM network's performance will be before the training and testing process is done. One of the factors that influences said process is the training dataset size used by the LSTM network. In order to assess the impact of such a factor on the end result, testing was carried out with varying training and testing dataset sizes. The case used in this study is the case of stock price prediction, where past stock prices are used to find out future stock prices. Four banks were chosen as the source of the initial dataset, and each initial dataset consists of six dataset features, namely: *Open*, *High*, *Low*, *Close*, *Adj. Close*, and *Volume*. Five variations of dataset sizes were used, that is: 70/30, 75/25, 80/20, 85/15, and 90/10. The results obtained from the testing show that the 90/10 dataset ratio provided the best *LSTM Network* performance for almost all dataset features.

Keywords: Training dataset size, LSTM network, Stock price, Bank.