# ABSTRACT

**DEDDY GUNAWAN TASLIM (01679210019)**

**A COMPARATIVE STUDY OF AUTOREGRESSIVE INTEGRATED MOVING AVERAGE AND LONG SHORT-TERM MEMORY IN FORECASTING VOLATILE TIME SERIES DATA**
(xii + 58 pages; 10 pictures; 4 tables)

Forecasting is an essential subject in economics and business. Autoregressive Integrated Moving Average (ARIMA) has been extensively used despite its weaknesses, from requiring a minimum number of data points to the assumed linearity of data, which are not always feasible in business data. Therefore, ARIMA is unable to predict well with volatile data. With recent advancements, the Long Short-Term Memory (LSTM) shows potential to address such weaknesses.

This research is aimed to identify the effect of the number of short and long data points of the time series data, as well as a model that is more suitable for handling volatile data due to missing values. Performance metrics used are the model accuracy measured with RMSE and the model run-time performance measured with the Python Timeit library.

This research concluded that LSTM (RMSE 0.037618) is more accurate than ARIMA (RMSE 0.062667) in a dataset of 60 data points and it is reverted in a longer dataset of 228 data points (RMSE of ARIMA 0.006949 to LSTM 0.036025). While in the case of missing values, LSTM outperforms ARIMA, although both models have decreased accuracy if the number of missing values is increased. In terms of run-time performance, ARIMA is significantly faster than LSTM.

Keywords: ARIMA, LSTM, time series, forecasting, volatile
References: 47 (1976 – 2022)

# ABSTRAK

**DEDDY GUNAWAN TASLIM (01679210019)**

**A COMPARATIVE STUDY OF AUTOREGRESSIVE INTEGRATED MOVING AVERAGE AND LONG SHORT-TERM MEMORY IN FORECASTING VOLATILE TIME SERIES DATA**

(xii + 58 halaman; 10 gambar; 4 tabel)

Peramalan adalah subjek penting dalam ekonomi dan bisnis. Autoregressive Integrated Moving Average (ARIMA) telah sering digunakan meskipun memiliki kelemahan, mulai dari membutuhkan jumlah titik data minimum hingga asumsi linearitas data, yang mana tidak selalu memungkinkan dalam data bisnis. Karenanya, ARIMA tidak dapat memprediksi dengan baik pada data bervolatilitas tinggi. Dengan kemajuan terkini, Long Short-Term Memory (LSTM) menunjukkan potensi untuk mengatasi kelemahan tersebut.

Penelitian ini bertujuan untuk mengidentifikasi pengaruh dari jumlah titik data deret waktu yang pendek dan panjang, serta model yang lebih sesuai untuk menangani data bervolatilitas tinggi akibat *missing value*. Metrik performa yang diukur adalah akurasi model menggunakan RMSE dan lama waktu permodelan diukur dengan perintah standar Python Timeit.

Penelitian ini menyimpulkan bahwa LSTM (RMSE 0.037618) lebih akurat daripada ARIMA (RMSE 0.062667) pada kumpulan data 60 titik data dan berkebalikan pada kumpulan data yang lebih panjang yaitu 228 titik data (RMSE ARIMA 0.006949 dan LSTM 0.036025). Sedangkan untuk kasus *missing value*, LSTM mengungguli ARIMA meskipun kedua model mengalami penurunan akurasi jika jumlah *missing value* bertambah. Dalam hal lama waktu permodelan, ARIMA jauh lebih cepat daripada LSTM.

Kata Kunci: ARIMA, LSTM, time series, forecasting, volatile
Referensi: 47 (1976 – 2022)