

## ABSTRAK

Kegiatan investasi saham mengalami peningkatan dari segi jumlah investor sebesar 53,41% pada era pandemi COVID-19, sehingga membuktikan bahwa kekuatan dunia pasar modal lebih diprioritaskan oleh masyarakat. Fakta tersebut yang akhirnya mendorong para investor untuk berlomba-lomba untuk mendapatkan informasi sebanyak-banyaknya untuk mendapatkan *return* yang besar dalam waktu singkat. Oleh karena itu, banyak investor berdiskusi, memprediksi, dan membagikan berita mengenai saham-saham melalui media sosial, salah satunya adalah Twitter. Melalui pesan teks Twitter tersebut, model analisis sentimen seperti FinBERT mampu berperan dalam menentukan polaritas dari teks tersebut. Tujuan dari penelitian ini adalah menentukan performa dari model SARIMAX dan LSTM dalam memprediksi perubahan harga saham Grup GOTO dengan membandingkan metrik RMSE dan MAE dari model tersebut dan mengetahui pengaruh variabel-variabel eksogen yang digunakan sebagai *input* dari model-model tersebut dalam memprediksi perubahan harga saham. Hasil yang didapatkan pada penelitian ini adalah model LSTM dengan variabel eksogen adalah model dengan performa terbaik dari model kontrol tanpa variabel eksogen, seperti LSTM *baseline* dan SARIMA, serta SARIMAX dalam memprediksi perubahan harga saham PT GoTo Gojek Tokopedia Tbk. berdasarkan nilai metrik RMSE (3,68) dan MAE (2,47). Model LSTM dengan eksogen memiliki peningkatan performa nilai RMSE sebesar 4% dan MAE sebesar 2% dibandingkan dengan model LSTM *baseline*. Variabel-variabel eksogen dalam model LSTM berhasil meningkatkan performa model, komponen sentimen dari *tweets* di Twitter (*positive\_adj*) yang diolah menggunakan model FinBERT memiliki pengaruh yang terbesar dibandingkan variabel eksogen yang lainnya, diikuti oleh jumlah *reply* harian dari *metadata tweets* Twitter. Model berbasis SARIMAX mengalami degradasi performa dengan penambahan variabel eksogen, hal tersebut dapat dilihat dari nilai metrik RMSE dan juga MAE.

Kata kunci: GOTO, Analisis Sentimen, LSTM, SARIMAX

## ABSTRACT

*The COVID-19 pandemic has caused an increase in the number of investors by 53.41%, thus cementing the capital market as a priority for many people. That fact causes investors to collect as much information as possible to gain enormous returns in short period of time. Investors discuss, predict, and share stories through social media platforms, such as Twitter. Through Twitter messages, one can perform sentiment analysis using models like FinBERT to find the polarity for those messages. The purposes of this research are determining the performance of the SARIMAX and LSTM models in predicting the price change for GOTO Group by comparing their RMSE and MAE and to understand the impact of exogenous variables in stock price change modelling. Based on the RMSE (3.68) and MAE (2.47) metric values, the LSTM model with exogenous variables outperforms the control model without exogenous variables, such as the LSTM baseline and SARIMA, and SARIMAX, in predicting changes in the stock price of PT GoTo Gojek Tokopedia Tbk. When compared to the baseline LSTM model, the LSTM model with exogenous has a 4% increase in RMSE and a 2% increase in MAE performance. The exogenous variables in the LSTM model have succeeded in improving the model's performance. The sentiment component of tweets on Twitter (positive\_adj) processed using the FinBERT model has the greatest influence compared to other exogenous variables, followed by the number of daily replies from Twitter tweet metadata. The addition of exogenous variables degrades the performance of the SARIMAX-based model, as evidenced by the RMSE and MAE metric values.*

*Keywords: GOTO, Sentiment Analysis, LSTM, SARIMAX.*