

INTISARI

Integrasi Klasterisasi dan LSTM dalam Optimalisasi Prediksi Harga Saham

(Studi Kasus : Saham LQ45 Periode Agustus 2024)

Oleh

MAWAR IDAH SHONIA

23/524953/PPA/06586

Prediksi harga saham merupakan tantangan signifikan dalam bidang keuangan karena karakteristik data yang kompleks dan bersifat volatil. Model prediktif tradisional seringkali gagal menangkap pola nonlinier yang tersembunyi dalam data, sementara efisiensi komputasi menjadi kendala tersendiri ketika menangani data berskala besar. Untuk menjawab permasalahan tersebut, penelitian ini mengintegrasikan pendekatan *hierarchical clustering* ke dalam proses prediksi harga saham dengan menggunakan dua model utama: *Long Short-Term Memory* (LSTM) dan *Autoregressive Integrated Moving Average* (ARIMA). Proses klasterisasi dilakukan terhadap data saham LQ45 menggunakan tiga fungsi jarak, yaitu Euclidean, Manhattan, dan Mahalanobis, untuk mengelompokkan saham berdasarkan kemiripan pola pergerakan *return* harga. Hasil klasterisasi menunjukkan bahwa klaster 1 cenderung memiliki pola pergerakan yang menurun dan lebih fluktuatif, sedangkan klaster 2 menunjukkan pola yang lebih stabil. Setiap klaster kemudian diprediksi secara terpisah guna meningkatkan efisiensi dan akurasi model. Evaluasi dilakukan menggunakan metrik MSE dan RMSE. Hasil evaluasi menunjukkan bahwa model LSTM dengan klasterisasi berbasis fungsi jarak Euclidean menghasilkan performa terbaik, dengan nilai MSE sebesar 0,0000995 dan RMSE sebesar 0,009923. Jika dibandingkan dengan model tanpa klasterisasi atau dengan model ARIMA, pendekatan LSTM berbasis klasterisasi mampu memberikan prediksi yang lebih akurat dan stabil, khususnya dalam menangkap dinamika harga saham yang bersifat nonlinier dan kompleks.

ABSTRACT

Integration of Clustering and LSTM in Optimizing Stock Price Prediction

(Case Study: LQ45 Stocks August 2024 Period)

By

MAWAR IDAH SHONIA

23/524953/PPA/06586

Stock price prediction remains a significant challenge in the field of finance due to the complex and volatile nature of financial data. Traditional predictive models often fail to capture the nonlinear patterns embedded in the data, while computational efficiency poses an additional constraint, particularly when dealing with large-scale datasets. To address these challenges, this study integrates a *hierarchical clustering* approach into the stock price prediction process using two primary models: *Long Short-Term Memory* (LSTM) and *Autoregressive Integrated Moving Average* (ARIMA). Clustering is applied to the LQ45 stock data using three distance functions—Euclidean, Manhattan, and Mahalanobis to group stocks based on the similarity of their return patterns. The clustering results reveal that Cluster 1 tends to exhibit a more volatile and declining trend, while Cluster 2 demonstrates greater stability. Each cluster is then modeled separately to enhance both computational efficiency and predictive accuracy. Model performance is evaluated using two key metrics: Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). The evaluation results indicate that the LSTM model, combined with clustering based on the Euclidean distance function, delivers the best performance, achieving an MSE of 0.0000995 and an RMSE of 0.009923. Compared to non-clustered models or those using ARIMA, the clustering-based LSTM approach provides more accurate and stable predictions, particularly in capturing the nonlinear and complex dynamics of stock price movements.