

Implementasi dan Evaluasi Metode Feature Engineering FinBERT untuk Prediksi Harga Saham Berbasis Sentimen Analisis Menggunakan BiLSTM = Implementation and Evaluation of FinBERT Feature Engineering Method for Sentiment Analysis-Based Stock Price Prediction Using BiLSTM

Mochamad Reza Rahadi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920526256&lokasi=lokal>

Abstrak

Meskipun teknologi telah mempengaruhi hampir semua aspek industri finansial, penelitian yang terfokus pada penggunaan teknologi pemrosesan teks dan analisis sentimen dalam konteks prediksi harga saham masih belum banyak dilakukan. Manfaat dan potensi dari penelitian semacam ini memiliki pengaruh yang tinggi, terutama karena analisis sentimen telah menjadi komponen yang penting dalam memprediksi tren pasar saham. Dalam penelitian ini, penulis mengusulkan penerapan metode feature engineering dalam memprediksi harga saham dengan memanfaatkan hasil analisis sentimen menggunakan FinBERT, lalu hasilnya akan dijadikan fitur oleh model BiLSTM. FinBERT adalah model berbasis BERT yang telah dilatih khusus untuk memproses dan menginterpretasi teks keuangan, sementara BiLSTM adalah arsitektur jaringan saraf berulang yang mampu mengatasi masalah yang ada pada jaringan saraf berulang standar seperti vanishing gradient dan efektif dalam mengolah data sekuensial. Penelitian ini menggabungkan kedua teknik ini untuk menciptakan model yang mampu memprediksi pergerakan harga saham berdasarkan analisis sentimen berita keuangan dengan nilai rata-rata MSE yang lebih rendah. Feature engineering digunakan dalam penelitian ini untuk mengekstrak dan mengolah informasi yang relevan dari dataset oleh model FinBERT untuk digunakan pada model BiLSTM. Dengan menggunakan metode feature engineering, ditemukan bahwa model BiLSTM yang menggunakan fitur sentimen analisis memiliki performa tertinggi dengan memiliki rata-rata nilai MSE terkecil dalam memprediksi tujuh saham yang memiliki karakteristik berbeda dengan nilai 3.43, nilai tersebut merupakan rata-rata terkecil dibandingkan tiga model lain dalam penelitian ini seperti LSTM dengan nilai MSE 4.04, Random Forest dengan nilai MSE 9.77, dan SVM dengan nilai MSE 12.56. Selanjutnya, proses optimisasi model BiLSTM menggunakan Optuna ditemukan nilai hyperparameter terbaik dalam menghadapi tujuh jenis saham yang berbeda, sehingga model mampu memprediksi lebih akurat dengan penurunan rata-rata nilai MSE hingga 40.55%. Sebagai bentuk validasi akhir pada penelitian ini telah dilakukan uji fold untuk mendapatkan model yang tidak overfitting dan memiliki rata-rata nilai MSE terkecil dengan variasi nilai hyperparameter batch size. Ditemukan batch size 16 merupakan ukuran paling optimal untuk tipe data NVDA, XOM dan TSLA dengan rata-rata MSE terkecil 0.64, 0.35, 0.05 sedangkan batch size 24 merupakan ukuran paling optimal untuk tipe data saham AAPL, AMZN, GOOG dan GOOGL dengan rata-rata MSE terkecil 0.028, 0.02, 0.03, 0.04, dan 0.03. Dalam menggunakan fitur sentimen analisis berhasil membuktikan menurunkan nilai MSE pada masing-masing jenis saham hingga rata-rata penurunan nilai MSE mencapai 33.10% dari semua jenis variasi data saham tanpa menggunakan fitur sentimen.

.....Although technology has influenced nearly all aspects of the financial industry, there is still a lack of research focusing on the use of text processing technology and sentiment analysis in the context of stock price prediction. The benefits and potential of such research are significant, especially as sentiment analysis

has become a crucial component in predicting stock market trends. In this study, the authors propose the application of feature engineering to predict stock prices by utilizing sentiment analysis results using FinBERT, which are then used as features by the BiLSTM model. FinBERT is a BERT-based model specifically trained to process and interpret financial text, while BiLSTM is a recurrent neural network architecture capable of overcoming problems inherent in standard recurrent neural networks, such as the vanishing gradient, and is effective in processing sequential data. This study combines these two techniques to create a model capable of predicting stock price movements based on sentiment analysis of financial news with a lower average MSE value. Feature engineering is used in this study to extract and process relevant information from the dataset by the FinBERT model to be used in the BiLSTM model. By using feature engineering, it was found that the BiLSTM model using sentiment analysis features has the highest performance, having the lowest average MSE value in predicting seven stocks with different characteristics, with a value of 3.43, which is the smallest average compared to the three other models in this study, such as LSTM with an MSE value of 4.04, Random Forest with an MSE value of 9.77, and SVM with an MSE value of 12.56. Furthermore, the optimization process of the BiLSTM model using Optuna found the best hyperparameters in dealing with seven different types of stocks, enabling the model to predict more accurately with an average reduction in MSE value up to 40.55%. As a final form of validation in this study, a fold test was conducted to obtain a model that is not overfitting and has the smallest average MSE value with variations in hyperparameter batch size values. It was found that a batch size of 16 is the most optimal size for NVDA, XOM, and TSLA data types with the smallest average MSE of 0.64, 0.35, 0.05, while a batch size of 24 is the most optimal size for AAPL, AMZN, GOOG, and GOOGL stock data types with the smallest average MSE of 0.028, 0.02, 0.03, 0.04, and 0.03. Using sentiment analysis features proved to reduce the MSE value for each type of stock to an average reduction in MSE value reaching 33.10% from all types of stock data variations without using sentiment features