ABSTRAK

Bitcoin sebagai aset kripto dengan kapitalisasi pasar terbesar mengalami volatilitas harga tinggi yang dipengaruhi oleh data historis pasar dan sentimen publik melalui media berita. Volatilitas ini menimbulkan risiko finansial bagi investor. Penelitian sebelumnya sering menggunakan data media sosial dengan model sentimen umum seperti VADER atau TextBlob, yang kurang menangkap konteks kripto spesifik dan rentan noise dari postingan manipulatif. Penelitian ini bertujuan membangun model prediksi harga Bitcoin menggunakan Long Short-Term Memory (LSTM) dengan integrasi analisis sentimen berbasis CryptoBERT untuk meningkatkan akurasi.

Data yang digunakan meliputi harga penutupan Bitcoin dari Yahoo Finance (1 Januari 2020–31 Desember 2025) dan berita dari Cointelegraph via web scraping serta dataset Hugging Face. Berita dianalisis dengan CryptoBERT milik ElKulako, menghasilkan sentimen Bullish, Neutral, dan Bearish yang dikonversi numerik dan diagregasi harian via rata-rata. Integrasi early fusion menggabungkan harga penutupan dan sentimen sebagai fitur input. Dataset dibagi dengan persentase 83,3% training (2020–2024) dan 16,7% testing (2025). Model LSTM multivariat dioptimasi via Grid Search dengan 540 kombinasi hyperparameter (window size, hidden size, number of layers, batch size, learning rate). Pelatihan menggunakan L1Loss dan early stopping patience 10 untuk cegah overfitting.

Hasil tuning menunjukkan konfigurasi optimal: window size 7, hidden size 64, layers 1, batch size 128, learning rate 0,01, dengan validation loss 0,0131. Evaluasi testing: RMSE 2.475,71 USD, MAE 1.915,16 USD, MAPE 1,89%. MAPE di bawah 2% menunjukkan akurasi tinggi meski volatilitas tinggi. Prediksi 1 Januari 2026: 86.687,19 USD. Penelitian membuktikan integrasi CryptoBERT dengan LSTM meningkatkan akurasi dibanding model berbasis harga saja. Kontribusi: model sentimen domain-spesifik, sumber berita stabil, dan dukungan pengurangan risiko investor.

Kata Kunci: CryptoBERT, LSTM, Analisis Sentimen, Prediksi Bitcoin, Deep Learning

ABSTRACT

Bitcoin, as the cryptocurrency with the largest market capitalization, experiences high price volatility influenced by historical market data and public sentiment through news media. This volatility poses significant financial risks to investors. Previous studies often used social media data with general sentiment models like VADER or TextBlob, which are less capable of capturing crypto-specific contexts and are susceptible to noise from manipulative posts. This study aims to build a Bitcoin price prediction model using Long Short-Term Memory (LSTM) integrated with sentiment analysis based on CryptoBERT to improve accuracy.

The data used includes Bitcoin closing prices from Yahoo Finance (January 1, 2020–December 31, 2025) and news from Cointelegraph via web scraping as well as Hugging Face dataset. News is analyzed using ElKulako's CryptoBERT, producing Bullish, Neutral, and Bearish sentiments converted to numeric values and aggregated daily via averaging. Early fusion integration combines closing prices and sentiment as input features. The dataset is divided chronologically: 83.3% training (2020–2024) and 16.7% testing (2025). The multivariate LSTM model is optimized via Grid Search with 540 hyperparameter combinations (window size, hidden size, number of layers, batch size, learning rate). Training uses L1Loss and early stopping with patience 10 to prevent overfitting.

Tuning results show the optimal configuration: window size 7, hidden size 64, layers 1, batch size 128, learning rate 0.01, with validation loss 0.0131. Testing evaluation: RMSE 2,475.71 USD, MAE 1,915.16 USD, MAPE 1.89%. MAPE below 2% indicates high accuracy despite high volatility. Prediction for January 1, 2026: 86,687.19 USD. This study proves that integrating CryptoBERT with LSTM improves accuracy compared to price-based models alone. Contributions: domain-specific sentiment model and stable news sources to support investor risk reduction.

Keywords: *CryptoBERT, LSTM, Sentiment Analysis, Bitcoin Prediction, Deep Learning*