

## STOCK PRICE PREDICTION FOR MATERIALS SECTOR USING CNN AND BI-LSTM ALGORITHM

**Annisa Desianty<sup>1</sup>, Widang Muttaqin<sup>1</sup>**

<sup>1</sup>Information Technology, Telkom University

*email:* annisadesianty@telkomuniversity.ac.id, widangmuttaqin@telkomuniversity.ac.id

**Abstract:** The materials sector is one of the stock markets sectors that attracts investors due to the high level of construction activity in Indonesia, which supports long-term growth. Stock price movements are influenced by various factors, requiring investors to determine the appropriate timing for buying, selling, or holding stocks. Therefore, this study aims to predict stock prices in the materials sector using a combination of CNN–BiLSTM algorithms. The research data were obtained from Yahoo Finance and processed through min–max normalization, data splitting, sliding window, model implementation, and evaluation stages. Testing was conducted on INTP and SMGR stocks with data split scenarios ranging from 60:40 to 90:10. The results show that CNN–BiLSTM performs best with a 90:10 data split, with minimum MSE and MAPE values of 0.000153 and 2.471% for INTP, and 0.000199 and 2.208% for SMGR, respectively. These findings indicate that increasing the proportion of training data improves the model's ability to learn historical patterns and produce more stable predictions.

**Keywords:** CNN-BILSTM; materials sector; stock

**Abstrak:** Sektor materials merupakan salah satu sektor saham yang diminati investor karena tingginya aktivitas pembangunan di Indonesia yang mendorong pertumbuhan jangka panjang. Pergerakan harga saham dipengaruhi oleh berbagai faktor sehingga investor perlu menentukan waktu transaksi yang tepat. Oleh karena itu, penelitian ini bertujuan memprediksi harga saham sektor materials menggunakan kombinasi algoritma CNN–BiLSTM. Data penelitian diperoleh dari Yahoo Finance dan diproses melalui tahapan normalisasi min–max, pembagian data, sliding window, implementasi model, serta evaluasi. Pengujian dilakukan pada saham INTP dan SMGR dengan skenario pembagian data 60:40 hingga 90:10. Hasil menunjukkan bahwa CNN–BiLSTM menghasilkan performa terbaik pada pembagian data 90:10, dengan nilai MSE dan MAPE minimum masing-masing sebesar 0.000153 dan 2.471% untuk INTP, serta 0.000199 dan 2.208% untuk SMGR. Temuan ini mengindikasikan bahwa peningkatan porsi data latih meningkatkan kemampuan model dalam mempelajari pola historis dan menghasilkan prediksi yang lebih stabil.

**Kata kunci:** CNN-BILSTM; saham; sektor materials

## INTRODUCTION

In this digital era, there are various investment options that can be an alternative in determining the allocation of funds or resources [1]. One well-

known form of investment is investing in the capital market. One of the investment instruments available in the capital market is shares [2].

Shares are one of the main instruments in the capital market that can

provide profits for investors through capital gains, namely the difference between the selling price and the purchase price of shares [3]. One of the sectors is the materials (raw materials) sector.

The materials sector consists of companies that sell products and services used by other industries as raw materials in the production of finished goods [4]. This sector is highly sought after by investors due to the high level of development activity in Indonesia, such as toll roads and industrial development, which has contributed to its growth. This makes the materials sector attractive to investors due to its long-term growth potential [5]. Several companies in the materials sector, such as PT Indocement Tunggal Prakarsa Tbk (INTP) and PT Semen Indonesia (Persero) Tbk (SMGR), are likely investors' choice of INTP and SMGR shares.

INTP recorded a notable increase in cement and clinker sales in 2024, indicating positive business performance [6]. Meanwhile, SMGR is one of the largest cement producers in Indonesia and Southeast Asia, with a dominant domestic market share and strong export capacity [7]. Stock price movements are influenced by many factors, upon closer inspection, the primary factor driving stock price movements is changes in supply and demand over time [8].

Many investors in the stock market still practice speculation like traders, with investment decisions influenced by rumors or other people's opinions that do not necessarily result in profits [9] [10]. To obtain information regarding future stock prices and reduce the risk of loss in investing, a prediction is needed that is able to estimate future stock price movements.

Related research was conducted by

Mushliha (2024) [11]. This paper discusses the stock price prediction of Islamic banks in Indonesia. This study proposes a method combining Convolutional Neural Network (CNN) - Bidirectional Long Short-Term Memory (BiLSTM). The data used in this study are the closing prices of Bank Syariah Indonesia (BSI), Bank Tabungan Pensiunan Negara Syariah (BTPN Syariah), and Bank Panin Dubai Syariah (PDSB) from January 2, 2020, to July 4, 2024. Testing the three stocks has MAPE values of 2.376%, 2.092%, and 0.629%, respectively.

Several other related studies such as those conducted by Diash et al (2025) [12]. This paper discusses Ethereum price prediction. Therefore, this study applies CNN-BiLSTM. The results yield a MAPE value of 2.8546% and an  $R^2$  of 0.9415. Further research was conducted by Samsi and Hidayar (2024) [13] which discusses the forecasting of the rupiah exchange rate against the Japanese yen. Therefore, the rupiah exchange rate against the Japanese yen was forecasted using CNN-BiLSTM. The results of the study produced high accuracy, as evidenced by a MAPE value of 0.588116%.

Based on the description above, this study will predict stock prices in the materials sector. The novelty of this study is that it uses data from the materials sector, which will be used from Yahoo Finance, using two stock price data from PT Indocement Tunggal Prakarsa Tbk (INTP) and PT Semen Indonesia (Persero) Tbk (SMGR).

Studies that specifically investigate the effectiveness of hybrid deep learning models on materials sector stock data, particularly in the Indonesian market, remain limited. This indicates a research gap in understanding whether

advanced deep learning architectures that perform well in highly volatile markets.

Stock price prediction in this study employs a hybrid CNN–BiLSTM model, where CNN extracts local temporal features from stock price data and BiLSTM captures long-term dependencies in sequential patterns [11]. Unlike traditional machine learning models or deep learning models with a single architecture, CNN-BiLSTM is more suitable for complex financial time series data, which contains local fluctuations and long-term trend components. Therefore, this architecture is considered appropriate for modeling raw material sector stock prices, which require long-term dependency learning [11]. The results are expected to support investors in making more informed investment decisions.

## METHOD

The dataset used comes from Yahoo Finance with a time span of January 1, 2023 to November 1, 2025. The data used is data from PT Indocement Tunggal Prakarsa Tbk (INTP), and PT Semen Indonesia (Persero) Tbk (SMGR). The variables used include open, high, low, close, and volume, following the study by Anisa (2024) [14]

Prior to modeling, the data were normalized using Min–Max scaling within the range [0,1] to improve numerical stability and accelerate the learning process. The normalization was performed using Equation (1) [15] [16] [17].

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Description:

$x'$  is the normalized result value

$x$  is actual data value to be normalized,

$x_{\min}$  is minimum value of actual data,  
 $x_{\max}$  is Maximum value of actual data.

Next, the dataset is divided into training and testing sets using four data split scenarios: 60:40, 70:30, 80:20, and 90:10. The training data are used to train the model, while the testing data are used to evaluate its performance [18] [19].

Sliding window created to minimize the approximation error in time-structured data, where the window and segment sizes are increased until a smaller error is achieved [20]. To convert the time series data into a supervised learning format, a sliding window approach is applied. In this method, a fixed number of historical observations are used as input features to predict the next time-step value.

During the model implementation stage, stock price prediction is performed using a hybrid CNN–BiLSTM architecture. CNN is employed to extract local temporal features from stock price data, while BiLSTM captures bidirectional long-term dependencies in the sequential features extracted by the CNN layers [11]. The model is implemented using Google Colaboratory, and the hyperparameter configurations used in this study are summarized in:

Table 1. Hyperparameter [11]

Parameter	Value
Convolution filters	32, 16
Kernel size	3
Activation function	ReLU
BiLSTM hidden units	64, 50 (layer 1); 32 (layer 2)
Window size	5, 10, 15, 20, 25, 30
Batch size	100, 200
Epochs	37, 71, 78, 40, 46, 68, 90, 114, 128, 370, 500, 562, 600

Parameter	Value
Learning rate	0.001
Optimizer	Adam
Loss function	MeanSquaredError

Model performance is evaluated using Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) by comparing predicted values with actual stock prices on the test dataset. MAPE measures prediction accuracy in percentage terms, while MSE quantifies the average squared error between predicted and actual values. The evaluation metrics are calculated using Equations (2) and (3) [21] [22].

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (2)$$

Where  $A_t$  is actual data,  $F_t$  = Simulation model data,  $n$  = amount of data.

$$MSE = \frac{1}{N} \sum (Y - \bar{Y})^2 \quad (3)$$

Description:

$N$  is amount of data

$Y$  is original data label

$\bar{Y}$  is Prediction Label.

## RESULTS AND DISCUSSION

This chapter evaluates several data

split proportions, where the 90%:10% split achieved the best performance and is therefore discussed further.

For the INTP dataset, hyperparameter tuning was conducted based on Table 1 across 26 experimental scenarios. The results show MSE values ranging from 0.002167 to 0.000153 and MAPE values between 2.47% and 4.14%, indicating improved accuracy with an increasing number of epochs. The lowest MSE was obtained using 600 epochs and a batch size of 200, while the lowest MAPE was achieved with 46 epochs and a batch size of 200.

Similarly, for the SMGR dataset, the evaluation results show MSE values ranging from 0.001568 to 0.000199 and MAPE values between 2.21% and 4.14%. The best MSE was achieved with 600 epochs and a batch size of 200, whereas the lowest MAPE was obtained with 40 epochs and a batch size of 200.

Overall, the results indicate that increasing the number of epochs generally improves prediction accuracy, although the optimal configuration differs depending on the evaluation metric used. From the explanation above, the explanation of the minimum MSE and MAPE values in the INTP and SMGR datasets can be seen in Table 2.

Table 2. History Minimum MSE and Overall MAPE

Dataset	60:40	70:30	80:20	90:10
INTP	Minimum MSE: History: 26 Epochs: 600 Batch Size: 200 MSE: 0.000216 MAPE: 4.535253 Minimum MAPE: History: 20 Epochs: 370	Minimum MSE: History: 26 Epochs: 600 Batch Size: 200 MSE: 0.000169 MAPE: 9.672288 Minimum MAPE: History: 11 Epochs: 68	Minimum MSE: History: 26 Epochs: 600 Batch Size: 200 MSE: 0.000241 MAPE: 4.047589 Minimum MAPE: History: 7 Epochs: 40 Batch Size: 100	Minimum MSE: History: 26 Epochs: 600 Batch Size: 200 MSE: 0.000153 MAPE: 3.763162 Minimum MAPE: History: 10 Epochs: 46

Dataset	60:40	70:30	80:20	90:10
	Batch Size: 200 MSE: 0.000583 MAPE: 4.292	Batch Size: 100 MSE: 0.000985 MAPE: 5.33187	MSE: 0.000976 MAPE: 2.917947	Batch Size: 200 MSE: 0.000615 MAPE: 2.471449
SMGR	Minimum MSE: History: 26 Epochs: 600 BatchSize: 200 MSE: 0.000234 MAPE: 31.93113 Minimum MAPE: History: 10 Epochs: 46 BatchSize: 200 MSE: 0.000807 MAPE: 3.329235	Minimum MSE: History: 26 Epochs: 600 BatchSize: 200 MSE: 0.000236 MAPE: 13.33023 Minimum MAPE: History: 2 Epochs: 37 BatchSize: 200 MSE: 0.001749 MAPE: 5.040519	Minimum MSE: History: 26 Epochs: 600 BatchSize: 200 MSE: 0.000207 MAPE: 6.050218 Minimum MAPE: History: 9 Epochs: 46 BatchSize: 100 MSE: 0.001377 MAPE: 4.172467	Minimum MSE: History: 26 Epochs: 600 BatchSize: 200 MSE: 0.000199 MAPE: 4.131106 Minimum MAPE: History: 8 Epochs: 40 BatchSize: 200 MSE: 0.00071 MAPE: 2.208131

Based on the test results in Table 2, the performance of the CNN-BiLSTM model on the INTP dataset shows improvement as the proportion of training data increases. The 90:10 data split yields the lowest MSE and MAPE values compared to other scenarios, indicating that the use of larger training data allows the model to learn stock price patterns more comprehensively, thus minimizing prediction errors. Variations in MSE and MAPE values across training configurations also indicate that model optimization is influenced by a combination of hyperparameters such as the number of epochs and batch size, resulting in a trade-off between absolute and relative accuracy.

Consistent results were also shown on the SMGR dataset, where a 90:10 data split provided the best performance with the lowest error value. This confirms that the CNN-BiLSTM model performs best when trained on a large amount of historical data. Overall, this experiment shows that the data split ratio significantly impacts prediction performance, and the 90:10 ratio is the

most effective configuration for generating stock price predictions. Based on the test results in Table 2, the performance of the CNN-BiLSTM model on the INTP dataset increases as the training data proportion increases, with a 90:10 data split producing the lowest MSE and MAPE values. This indicates that using a larger training data set allows the model to learn stock price patterns more comprehensively, thus minimizing prediction errors. Differences in MSE and MAPE values across various configurations also indicate the influence of hyperparameter combinations, such as the number of epochs and batch size, on model optimization.

The consistent reduction in MSE and MAPE as the training data proportion increases indicates that the CNN-BiLSTM model benefits significantly from larger historical datasets. This behavior suggests that the convolutional layers are able to extract more stable local temporal patterns when sufficient data are available, while the BiLSTM component effectively captures

long-term dependencies in stock price movements. With larger training data, the model learns its parameters more effectively, leading to more stable and accurate predictions across time steps. In contrast, using a smaller training proportion limits the model's exposure to diverse market conditions, which increases the risk of generalization errors.

Differences in error values between the INTP and SMGR datasets may also be influenced by variations in stock price volatility and trading volume characteristics. SMGR exhibits higher error values in several configurations, indicating more complex price dynamics that are harder to model. Higher volatility increases the variance of prediction errors, which disproportionately affects MSE compared to MAPE, explaining the observed metric differences between datasets. This suggests that stock-specific characteristics play a role in determining the effectiveness of deep learning-based prediction models.

The following (Images 1 and Image 2) show the actual data and best predictions from each of the INTP and SMGR datasets.



Image 1. Actual Data and Predictions for INTP Shares

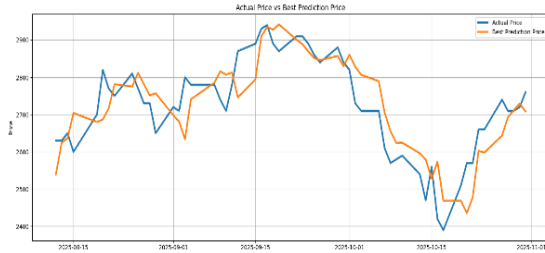


Image 2. Actual Data and Predictions for SMGR Shares

Comparatively, based on Image 1 and Image 2, the prediction model shows more stable performance for INTP shares than SMGR shares. This is indicated by the relatively smaller distance between the actual and predicted curves for INTP. SMGR shares have higher price fluctuations, resulting in a higher prediction error rate.

## CONCLUSION

This research successfully applied a combination of CNN-BiLSTM models to predict the stock prices of PT Indocement Tungal Prakarsa Tbk (INTP) and PT Semen Indonesia (Persero) Tbk (SMGR). Based on the test results with several data division proportions (60:40, 70:30, 80:20, and 90:10), the model showed good performance with low Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) values, thus having a high level of accuracy. Models with a larger number of epochs tend to produce the smallest MSE values, indicating the model's ability to learn stock data patterns well and provide stable prediction results.

In future research, it is recommended to use data from a longer time span or increase the variety of stock sectors to achieve more accurate and comprehensive predictions. Furthermore, the addition of external factors such as market conditions or economic data can be considered to enrich the analytical context and improve the accuracy of the predictions.

## BIBLIOGRAPHY

- [1] D. Saputro and D. Swanjaya,

- “Analisa Prediksi Harga Saham Menggunakan Neural Network Dan Net Foreign Flow,” *Gener. J.*, vol. 7, no. 2, pp. 96–104, 2023, doi: 10.29407/gj.v7i2.20001.
- [2] P. C. Hartono and A. D. Widianoro, “Analisis Prediksi Harga Saham Unilever Menggunakan Regresi Linier dengan RapidMiner,” *J. Comput. Inf. Syst. Ampera*, vol. 5, no. 3, pp. 2775–2496, 2024, [Online]. Available: <https://journal-computing.org/index.php/journal-cisa/index>
- [3] L. Alpianto, A. Hermawan, and Junaedi, “Moving Average untuk Prediksi Harga Saham dengan Linear Regression,” *J. Buana Inform.*, vol. 14, no. 02, pp. 117–126, 2023, doi: 10.24002/jbi.v14i02.7446.
- [4] D. Gunawan and T. Ferryanto, “Pengaruh Profitabilitas dan Solvabilitas Terhadap Harga Saham Sektor Bahan Baku Di Bursa Efek Indonesia,” *Account. Prog.*, vol. 3, no. 1, pp. 10–22, 2024.
- [5] F. A. Safitra, B. Wahyudiono, and F. Syafaat, “Pengaruh Profitabilitas, Pertumbuhan Penjualan Dan Ukuran Perusahaan Terhadap Harga Saham Perusahaan Sektor Bahan Baku Industri Logam Dan Mineral Yang Terdaftar Di Bursa Efek Indonesia Tahun 2019-2023,” *J. Manaj. Pratama*, vol. 2, no. 3, pp. 1–23, 2025.
- [6] C. A. Muchlis, “Begini Rekomendasi Saham Indocement Tunggal Prakarsa (INTP), Pangsa Pasar Naik,” *msn*. <https://www.msn.com/id-id/ekonomi/pasarpasar/begini-rekomendasi-saham-indocement-tunggal-prakarsa-intp-pangsa-pasar-naik/ar-AA1yqkou>
- [7] BRIDS, “Saham SMGR: Potensi Emiten Semen dalam Pembangunan Nasional,” *Brightsid*, 2025. <https://www.brights.id/id/blog/saham-smgr>
- [8] R. S. Santoso and F. K. Sari Dewi, “Konfigurasi Model Prophet Untuk Prediksi Harga Saham Sektor Teknologi di Indonesia Yang Akurat,” *J. Buana Inform.*, vol. 15, no. 01, pp. 50–58, 2024, doi: 10.24002/jbi.v15i1.8634.
- [9] A. D. B. Purba, D. A. Aruan, and F. Situmorang, “Analisis Faktor-Faktor yang Mempengaruhi Keputusan Investor dalam Memilih Saham di Pasar Modal Indonesia,” *AKUA J. Akuntansi dan Keuang.*, vol. 4, no. 3, pp. 473–481, 2025, doi: 10.54259/akua.v4i3.5160.
- [10] A. O. Pratama, K. Purba, J. Jamhur, and B. T. P. Pamungkas, “Pengaruh Faktor Perilaku Investor Saham Terhadap Keputusan Investasi di Bursa Efek Indonesia,” *Monet. J. Akunt. dan Keuang.*, vol. 7, no. 2, pp. 170–179, 2020.
- [11] M. Mushliha, “Implementasi CNN-BiLSTM untuk Prediksi Harga Saham Bank Syariah di Indonesia,” *Jambura J. Math.*, vol. 6, no. 2, pp. 195–203, 2024, doi: 10.37905/jjom.v6i2.26509.
- [12] H. D. Diash, V. Nathania, M. Idhom, and T. Trimono, “Application of CNN-BiLSTM Algorithm for Ethereum Price Prediction,” *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1709–1714, 2025, doi: 10.30871/jaic.v9i4.9757.
- [13] H. L. P. Samsi and Y. Hidayat, “Peramalan Nilai Tukar Rupiah Terhadap Yen Jepang Menggunakan Convolutional Neural Network Bidirectional Long Short Term Memory,” *Pros. Semin. Nas. Stat.*, vol. 11, pp. 7–19, 2024, [Online]. Available:

- <https://prosidings.statistics.unpad.ac.id/?journal=prosidingns&page=article&op=view&path%5B%5D=318>
- [14] Y. Anisa, M. Hafiz, and N. Novita, "Pengembangan Model Prediksi Harga Saham Dengan Menggunakan Regresi Linear Berganda Pada Saham BRI," *JISTech (Journal Islam. Sci. Technol.*, vol. 9, no. December, pp. 191–195, 2024.
- [15] N. P. S. Y. Artini, I. W. Sumarjaya, and D. P. E. Nilakusmawati, "Penerapan Metode Support Vector Regression (SVR) Dengan Penerapan Metode Support Vector Regression (Svr) Dengan Algoritma Grid Search Dalam," *E-Jurnal Mat.*, no. May, 2024, doi: 10.24843/MTK.2024.v13i02.p447.
- [16] M. Andriyani, S. Nurwilda, D. Zatusiva Haq, and D. Candra Rini Novitasari, "Prediksi Harga Beras Premium Tahun 2024 Menggunakan Metode Gradient Boosted Trees Regression," *J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.*, vol. 18, no. 2, pp. 75–84, 2024, [Online]. Available: <https://doi.org/10.47111/JTIAvailabl>  
[eonlineathttps://e-journal.upr.ac.id/index.php/JTI](https://e-journal.upr.ac.id/index.php/JTI)
- [17] N. Putu and N. Kusuma, "PREDIKSI HARGA SAHAM BLUE CHIP PADA INDEKS IDX30 MENGGUNAKAN ALGORITMA RECURRENT NEURAL NETWORK (RNN)," *J. Ekon. BISNIS*, vol. 23, no. 1, pp. 90–97, 2024.
- [18] M. N. Suryanooradja, A. D. Rahajoe, and A. Junaidi, "Hybrid Mobilenetv2 Dan Extreme Gradient Boosting Untuk Klasifikasi Kerusakan Bangunan," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 3, pp. 402–412, 2025, doi: 10.23960/jitet.v13i3.6858.
- [19] M. G. B. Ashari, "Implementasi Algoritma Convolutional Neural Network untuk Meningkatkan Identifikasi Penyakit Tanaman Durian," *Jupiter Publ. Ilmu Keteknikan Ind. Tek. Elektro dan Inform.*, vol. 2, no. 4, pp. 162–172, 2024, doi: 10.61132/jupiter.v2i4.418.
- [20] A. Rahmawati, W. Sulandari, S. Subanti, and Y. Yudhanto, "Penerapan Metode Reccurent Neural Network dengan Pendekatan Long Short-Term Memory (LSTM) untuk Meramalkan Harga Saham Hybe Corporation The Application of Recurrent Neural Network Method with the Long Short-Term Memory (LSTM) Approach to Forecast Hybe Cor," *J. Bumigora Inf. Technol.*, vol. 5, no. 1, pp. 65–76, 2023, doi: 10.30812/bite/v5i1.2973.
- [21] T. Han, K. Muhammad, T. Hussain, J. Lloret, and S. W. Baik, "An Efficient Deep Learning Framework for Intelligent Energy Management in IoT Networks," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3170–3179, 2021, doi: 10.1109/JIOT.2020.3013306.
- [22] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.