ISSN: 2807-324X (*Online*)

PREDIKSI HARGA SAHAM BERBASIS ANALISIS SENTIMEN MEDIA SOSIAL DAN *BIDIRECTIONAL* LSTM

Abduh Riski*, Abdul Mujib Shabirin, Alfian Futuhul Hadi, Ahmad Kamsyakawuni

Universitas Jember, Jalan Kalimantan No. 37 – Kampus Bumi Tegalboto, Jember, Jawa Timur, 68121, Indonesia
*riski.fmipa@unej.ac.id

Abstract. Stocks are a popular investment instrument due to their potential for significant returns. A company's stock prices fluctuate due to various factors, including market sentiment and investor opinions. Social media sentiment analysis enables the identification of public opinions relevant to a company. This study aims to predict the stock prices of PT Indofood CBP Sukses Makmur Tbk by utilizing sentiment analysis from social media and the Bidirectional LSTM (BILSTM) method. BILSTM is an extension of LSTM that can capture information in both directions, enhancing prediction accuracy. The study uses daily stock price data ($t-1$) and sentiment data from social media, divided into 90% for training (327 data points) and 10% for testing (37 data points). The best model architecture consists of 50 neurons, batch size 32, and 500 epochs. This model achieved a Mean Squared Error (MSE) of 16,621.26 and a Mean Absolute Percentage Error (MAPE) of 0.98%. The predicted stock closing price on May 31, 2023, was IDR 11,108.205. These results demonstrate that combining social media sentiment analysis and BILSTM effectively predicts stock prices for companies in the Fast Moving Consumer Goods (FMCG) sector.
Keywords: BILSTM, Prediction, Sentiment Analysis, Stocks

1. Pendahuluan

Pesatnya pertumbuhan ekonomi suatu negara membutuhkan pendanaan yang berasal dari berbagai pihak, baik masyarakat maupun pemerintah. Pasar modal menjadi salah satu sarana pengumpulan dana yang efektif untuk sektor swasta dan pemerintah. Melalui pasar modal, perusahaan dapat menerbitkan efek seperti obligasi dan saham yang dapat diperdagangkan. Saham merupakan bukti kepemilikan terhadap suatu perusahaan dan menjadi salah satu instrumen investasi yang digemari karena potensinya memberikan keuntungan yang besar.

Namun, usaha mendapatkan keuntungan dari saham tidaklah mudah. Investor harus mampu memprediksi harga saham agar dapat menentukan waktu yang tepat untuk membeli atau menjual. Selain itu, mereka juga berusaha meminimalkan risiko demi menjaga keamanan modal yang diinvestasikan. Untuk itu, diperlukan teknik analisis yang mendalam dalam memprediksi pergerakan harga saham.
Prediksi harga saham merupakan tantangan yang kompleks karena dipengaruhi oleh berbagai faktor, seperti kondisi fundamental perusahaan, kinerja sektor industri, dan sentimen publik. Sentimen publik, terutama dari media sosial, semakin relevan untuk dianalisis karena mencerminkan opini investor dan pasar secara *real-time*.

Beberapa penelitian telah menggunakan berbagai metode untuk memprediksi harga saham. Misalnya, penelitian Bhuriya dkk. (2017) menggunakan *deep learning* dengan metode *linear regression*, *radial basis functions* (RBF), dan *polynomial,* namun hasil akurasi berbeda-beda tergantung *dataset* yang digunakan. Penelitian lainnya, seperti Jahan & Sajal (2018), memanfaatkan *Recurrent Neural Network* (RNN) dan memperoleh nilai MSE yang cukup rendah. Pendekatan berbasis *Long Short-Term Memory* (LSTM) juga telah dikembangkan, seperti penelitian Mathur dkk. (2019) yang menggunakan LSTMP, DLSTMP, dan DLSTM, menunjukkan keunggulan metode ini dalam menangkap pola data sekuensial.

ISSN: 2807-324X (*Online*)

Sementara itu, penelitian Afrianto dkk. (2022) menggabungkan analisis sentimen dari media sosial dan *Bidirectional* LSTM (BiLSTM) untuk memprediksi harga saham PT Bank Central Asia Tbk. Hasil penelitian menunjukkan bahwa data sentimen dapat meningkatkan akurasi prediksi.

Dalam penelitian ini, dilakukan analisis sentimen dari media sosial Twitter dan metode BiLSTM untuk memprediksi harga saham PT Indofood CBP Sukses Makmur Tbk, salah satu perusahaan *Fast Moving Consumer Goods* (FMCG) terbesar di Indonesia. BiLSTM dipilih karena kemampuannya menangkap informasi dari dua arah, sehingga meningkatkan performa model dalam pengklasifikasian. Penulis juga menambahkan parameter timestep untuk memanfaatkan data historis dalam prediksi. Penelitian ini diharapkan memberikan kontribusi dalam pengembangan model prediksi harga saham berbasis sentimen publik secara lebih akurat dan relevan.

2. Metode

Data pada penelitian ini menggunakan 2 data sekunder yaitu data historis harga saham PT Indofood CBP Sukses Makmur Tbk yang diperoleh pada laman *finance.yahoo.com*. Data sentimen publik diperoleh dari algoritma *tweet harvest*. Periode data saham yang diambil mulai dari 1 Desember 2021 hingga 30 Mei 2023. Setelah data didapatkan akan dilakukan *preprocessing* data pada saham berupa *selection data*. *Preprocessing data* pada data sentimen mencakup *case folding, stopword removal, stemming* dan penyusunan kata. Kata yang telah disusun akan dilakukan perhitungan menggunakan algoritma VADER untuk menghasilkan nilai *compound*. Data sentimen dan saham akan digabungkan dan dilakukan pembuatan model. Model dengan nilai error terkecil akan menjadi model terbaik.

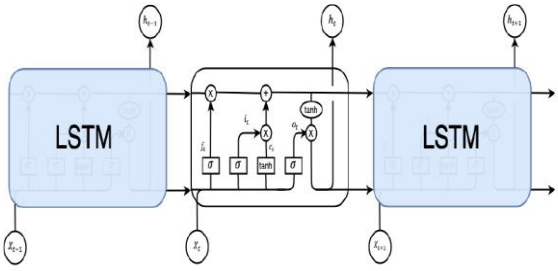
a. Valence Aware Dictionary and Sentiment Reasoner (VADER)

VADER merupakan pustaka dengan kemampuannya dalam mengakses sentimen pada tiap teks. Eric dan C.J Hutton merupakan orang yang pertama kali memperkenalkan VADER. VADER merupakan algoritma dengan pendekatan berbasis leksikon yaitu metode analisis sentimen yang menggunakan kamus leksikon berisi daftar kata berbagai opini (Nafan & Amalia, 2019).

Penilaian menggunakan VADER mengelompokkan kalimat menjadi 4 yaitu negatif, netral, positif dan *compound*. Satuan yang digunakan untuk pembagian kalimat menggunakan *compound*. Nilai *compound* ditentukan dari pengurutan kata, penguat bentuk kata, penguat tanda baca dan skor valensi dari setiap kata akan dijumlahkan. Terakhir, skor akan mengalami normalisasi dengan rentang -1 sampai dengan +1. Sentimen positif memiliki nilai compound ≥ +0,05, sentimen negatif memiliki nilai *compound* ≤ -0,05 dan Sentimen netral memiliki nilai compound antara –0,05 dan +0,05 (Karim & Das, 2018).

b. Long Short Term Memory (LSTM)

Schmidhuber dan Sepp Hochreiter adalah orang yang pertama kali menjelaskan LSTM pada tahun 1997. Dengan adanya, *memory* cell LSTM yang dapat bertambah maka mampu menyelesaikan masalah RNN berupa terbentuknya *vanishing gradient* disaat menjalankan data berurutan yang panjang. *Forget gate*, *input gate*, dan *output gate* merupakan *gates* pada LSTM. Selain itu, *tanh* dan *sigmoid* (σ) sebagai fungsi aktivasinya (Westergaard dkk., 2018). Arsitektur LSTM ditunjukkan pada Gambar 1.



Gambar 1. Arsitektur LSTM

ISSN: 2807-324X (*Online*)

Proses pembentukan model dibagi menjadi beberapa tahapan berikut :

Tahapan pertama yaitu menghitung forget gate (f_t) menggunakan persamaan (2),

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \quad (2)$$

Berikutnya, melakukan perhitungan input gate (i_t) menggunakan persamaan (3),

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (3)$$

Selanjutnya, perhitungan *update cell state* menggunakan persamaan (5) dengan menggunakan kandidat *cell state* ($C_{\bar{t}}$) pada persamaan (4),

$$C_{\bar{t}} = tanh\ (W_c.[h_{t-1}, x_t] + b_c) \quad (4)$$

$$C_t = f_t\ *\ C_{t-1} + i_t\ *\ C_{\bar{t}} \quad (5)$$

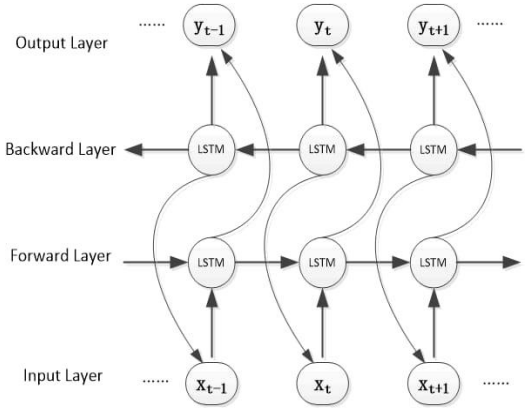
Terakhir, perhitungan ouput gate (o_t) menggunakan persamaan (6) dan menghasilkan hasil keluaran LSTM waktu t (h_t) seperti pada persamaan (7),

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t =\ o_t\ *\ tanh(C_t) \quad (7)$$

c. Bidirectional LSTM (BILSTM)

Bidirectional *LSTM* merupakan variasi dari metode LSTM yang dirancang oleh Paliwal dan Schuster sebagai pelatihan model jaringan saraf dengan data berurutan. *Input forward* dan *input backward* digunakan sebagai *input* dari model BiLSTM. Keluaran dari proses ini akan bersatu. Berdasarkan modelnya, data masa lampau dan masa depan pada tiap data berurutan dapat dipelajari. Arsitektur BiLSTM dapat dilihat pada Gambar 2.



Gambar 2. Arsitektur BILSTM

Perhitungan output dua hidden layer dapat menggunakan persamaan (8)

$$y_t = w_{\vec{h}y}\vec{h}_t + w_{\overleftarrow{h}y}\overleftarrow{h}_t \quad (8)$$

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Data yang digunakan pada penelitian ini yaitu data sentimen yang berasal dari media sosial twitter dan data historis harga saham PT Indofood CBP Sukses Makmur Tbk. Data twitter diambil menggunakan algoritma *tweet harvest* dan harga saham diambil melalui laman *yahoo finance* dengan periode tanggal 01 Desember 2021 hingga 31 Mei 2023. Plot data harga penutupan saham ditunjukkan pada Gambar 3.

ISSN: 2807-324X (*Online*)

Gambar 3. Plot harga penutupan saham

Berdasarkan Gambar 3 harga penutupan saham PT Indofood CBP Sukses Makmur Tbk mengalami kenaikan. Titik terendah ditunjukkan pada tanggal 18 April 2022 dengan harga saham Rp 7.100,00 dan titik tertingginya diperoleh pada tanggal 24 Mei 2023 dengan harga saham Rp 11.200,00.

3.2. Preprocessing Data

Data tweet yang menjadi data sentimen publik diambil pada tanggal yang sama pada hari perdagangan sedang berlangsung. Contoh data teks yang didapatkan adalah "Makin Gurih! Rekomendasi Beli Saham Indofood CBP (ICBP), TP Rp11.000 " dan “Tren Meningkat, Ekspor Mie Instan Indonesia Tembus Pasar Non-Tradisional”. Setelah data sentimen diproses akan dilakukan pemeriksaan nilai korelasi antara *nilai compound* dan harga penutupan. Korelasi yang ditunjukkan antara harga penutupan saham dan nilai *compound* adalah 0,0367. Setelah melakukan korelasi, data saham dan nilai compound akan digabungkan dan dinormalisasi. Variabel yang digunakan yaitu *Open, High, Low, Close,* dan *Compound*. Normalisasi data dilakukan untuk merubah data yang tersedia dengan rentang nilai antara 0 dan 1. Tabel 1 menunujukkan hasil dari normalisasi data

Tabel 1. Hasil Normalisasi

Date	***Open***	***High***	***Low***	***Close***	***Compound***
2021-12-01	0,393	0,418	0,4	0,445	0,715
2021-12-02	0,435	0,424	0,424	0,426	0,811
2021-12-03	0,417	0,393	0,393	0,390	0,240
…	…	…	…	…	…
2023-05-26	1,0	0,975	1,0	1,0	0,668
2023-05-29	1,0	0,975	0,993	0,993	0,832
2023-05-30	1,0	0,993	1,0	1,0	0,430

3.3. Pembagian Data

Data yang tersedia akan dibagi menjadi data *training* dan data *testing*. Pembagian data dibagi menjadi 3 percobaan dengan perbandingan data *training* dan *testing* yaitu 75%:25%, 80%:20%, dan 90%:10%. Model yang telah diperoleh, selanjutnya akan dilakukan pengujian terhadap data *testing* untuk melihat performa model yang telah terbentuk.

3.4. Pengujian *Model*

Setelah melakukan training data, tahap selanjutnya yaitu pengujian terhadap model. Pengujian dilakukan untuk menguji model yang telah dilatih dalam memprediksi harga saham. Pengujian model dilakukan dengan beberapa percobaan pada pembagian data, *timestep* dan penambahan variabel sentimen melalui nilai *compound*. *Timestep* yang dilakukan dalam penelitian ini adalah $t-1, t-2, t-3, t-4,$ dan $t-5$. Tidak ada aturan khusus dalam penentuan parameter, untuk mencapai hasil

ISSN: 2807-324X (*Online*)

terbaik perlu dilakukan beberapa percobaan pada parameter yang digunakan dalam membangun model. Tabel 2 merupakan hasil pengujian model tanpa melibatkan data sentimen.

Tabel 2. Hasil Pengujian Model Tanpa Melibatkan Data Sentimen

Perbandingan Data	*Timestep*	MAPE	MSE
75% : 25%	$t-1$	1.08%	18.658,90
	$t-2$	1.09%	18.776,35
	$t-3$	1.12%	21.719,73
	$t-4$	1.44%	37.155,25
	$t-5$	1.47%	38.180,55
80% : 20%	$t-1$	1.07%	20.731,41
	$t-2$	1.18%	25.176,99
	$t-3$	1.24%	28.847,27
	$t-4$	1.32%	31.883,13
	$t-5$	1.46%	38.526,93
90% : 10%	**$t-1$**	**1.02%**	**16.681,44**
	$t-2$	1.07%	17.690,66
	$t-3$	1.23%	22.515,22
	$t-4$	1.43%	31.876,33
	$t-5$	1.74%	52.026,04

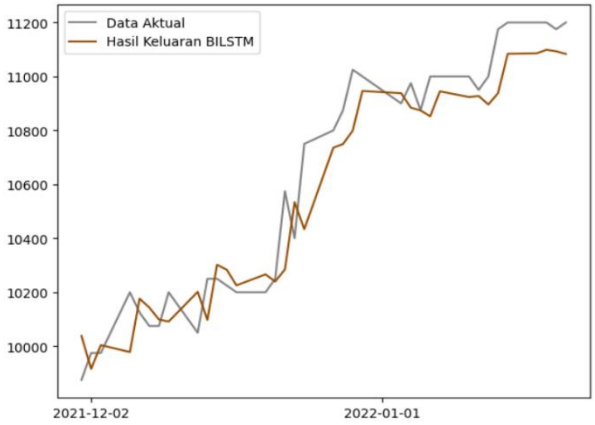
Berdasarkan Tabel 2 ditunjukkan bahwa semakin banyak dalam penggunaan jumlah timestep maka akan memberikan hasil yang buruk daripada menggunakan timestep yang lebih sedikit. Dari Tabel 3 diperoleh hasil terbaik yaitu penggunaan timestep $t-1$ dengan pembagian data training sebesar 90% dan data testing sebesar 10%. Pembagian data tesebut memuat 327 data training dan 37 data testing. Nilai MAPE yang diperoleh sebesar 1,02%, nilai ini termasuk kecil dikarenakan nilai MAPE dibawah 10% termasuk kategori sangat baik. Nilai MSE sebesar 16.681,44, nilai ini termasuk kecil untuk kesalahan rata - rata kuadrat dibandingkan skala yang digunakan dalam data yaitu sekitar 7.000 hingga 10.000. Tabel 3 menunjukkan hasil pengujian model menggunakan analisis sentimen.

Tabel 3. Hasil Pengujian Model Menggunakan Data Sentimen

Perbandingan Data	*Timestep*	MAPE	MSE
75% : 25%	$t-1$	1.00%	17.602,69
	$t-2$	1.05%	18.090,61
	$t-3$	1.10%	20.575,91
	$t-4$	1.32%	26.507,90
	$t-5$	1.35%	32.561,73
80% : 20%	$t-1$	1.05%	20.235,86
	$t-2$	1.16%	25.000,37
	$t-3$	1.17%	21.085,86
	$t-4$	1.18%	25.031,20
	$t-5$	1.22%	27.161,53
90% : 10%	**$t-1$**	**0.98%**	**16.621,26**
	$t-2$	1.06%	17.286,10
	$t-3$	1.16%	20.440,10
	$t-4$	1.20%	21.004,44
	$t-5$	1.72%	50.563,19

ISSN: 2807-324X (*Online*)

Berdasarkan Tabel 3 dapat ditunjukkan bahwa hasil terbaik yang diperoleh yaitu penggunaan timestep $t-1$ dengan pembagian data training sebesar 90% dan data testing sebesar 10%. Penggunaan parameter tebaik sama sepeti tanpa melibatkan data sentimen. Nilai MAPE yang diperoleh yaitu 0.98% dan nilai MSE yaitu 16.621,26. Hasil pengujian model tanpa menggunakan analisis sentimen memiliki nilai error lebih besar daripada model yang melibatkan variabel sentimen baik menggunakan MAPE maupun MSE. Sehingga, model terbaiknya adalah BiLSTM menggunakan analisis sentimen dengan parameter timestep $(t-1)$ dan pembagian data training sebesar 90% dan data testing sebesar 10%. Hasil ini juga didukung dari penelitian Renault (2017) yang menyatakan bahwa sentimen investor dapat berpengaruh pada prediksi pola indeks saham. Gambar 4 menunjukkan plot data aktual pada data testing dan hasil peramalan penutupan harga saham menggunakan metode BILSTM dan data sentimen.



Gambar 4. Plot keluaran model terbaik dan data aktual

Gambar 4 menunjukkan plot hasil peramalan dari BiLSTM menggunakan data sentimen dan data aktual menggunakan data testing sebanyak 36 data.Garis berwarna abu – abu menunjukkan data aktual pada data testing dan garis coklat menunjukkan hasil peramalan menggunakan BILSTM menggunakan data sentimen. Plot di atas menunjukkan bahwa hasil keluaran yang didapatkan dari data *testing* dapat mengikuti pola dari data aktual. Hal tersebut menunjukkan bahwa model dapat mempelajari pola dari data testing yang ada. Hasil keluaran BILSTM menggunakan data sentimen dilakukan denormalisasi. Data yang sudah didenormalisasi akan dievaluasi dengan data *testing* menggunakan MAPE dan MSE. Nilai MAPE dan MSE yang didapatkan yaitu 0.98% dan 16.621,26. Tabel 4 menunjukkan hasil keluaran BILSTM menggunakan data sentimen.

Tabel 4. Perbandingan Data Aktual dan *Output* Model Terbaik

Tanggal	**Data Aktual**	***Output***
2021-12-01	9.875	10.038,163
2021-12-02	9.975	9.916,121
2021-12-03	9.975	10.004,162
…	…	…
2023-05-28	11.200	11.098,784
2023-05-29	11.175	11.093,245
2023-05-30	11.200	11.083,206
2023-05-31		11.108,205

Pada tanggal 31 Mei 2023 yang ditunjukkan pada Tabel 4 menunjukkan hasil prediksi penutupan harga saham yaitu Rp 11.108,205. Prediksi dilakukan menggunakan data satu hari sebelumnya $(t-1)$ yaitu tanggal 30 Mei 2023 dalam memprediksi satu hari berikutnya.

ISSN: 2807-324X (*Online*)

4. Penutup

Berdasarkan hasil penelitian yang telah dilakukan diperoleh bahwa model terbaik dalam meramalkan penutupan harga saham yaitu menggunakan parameter timestep $t - 1$ dengan data training sebesar 90% dan data testing sebesar 10%. Hasil pengujian model yang diperoleh yaitu nilai MAPE sebesar 0.98% dan MSE sebesar 16.621,26. Nilai MAPE dibawah 10% termasuk kategori prediksi sangat baik dan nilai MSE yang diperoleh termasuk kategori prediksi sangat baik dan nilai MSE yang diperoleh termasuk kecil dari rentang data 7.000 sampai 12.000. Hasil prediksi penutupan harga saham pada tanggal 31 Mei 2023 dari model terbaik menggunakan metode analisis sentimen dan Bidirectional LSTM (BILSTM) adalah Rp 11.108,205. Hasil prediksi menunjukkan terjadinya penurunan dibandingkan harga penutupan saham pada tanggal 30 Mei 2023 yaitu Rp 11.200,-.

Daftar Pustaka

Afrianto, N., Fudholi, D. H.,dan Rani, S. (2022). Prediksi Harga Saham Menggunakan BiLSTM dengan Faktor Sentimen Publik. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *6*(1), 41–46. https://doi.org/10.29207/resti.v6i1.3676

Bhuriya, D., Kaushal, G., Sharma, A., dan Singh, U. (2017). Stock market predication using a linear regression. *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, 510–513. https://doi.org/10.1109/ICECA.2017.8212716

Jahan, I., dan Sajal, S. (2018). Stock price prediction using recurrent neural network (RNN) algorithm on time-series data. *Midwest instruction and computing symposium*.

Karim, M., dan Das, S. (2018). Sentiment Analysis on Textual Reviews. *IOP Conference Series: Materials Science and Engineering*, *396*, 012020. https://doi.org/10.1088/1757-899X/396/1/012020

Mathur, R., Pathak, V., dan Bandil, D. (2019). *Parkinson Disease Prediction Using Machine Learning Algorithm* (hlm. 357–363). https://doi.org/10.1007/978-981-13-2285-3_42

Nafan, M. Z., dan Amalia, A. E. (2019). Kecenderungan Tanggapan Masyarakat terhadap Ekonomi Indonesia berbasis Lexicon Based Sentiment Analysis. *Jurnal Media Informatika Budidarma*, *3*(4), 268–273.

Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance*, *84*, 25–40. https://doi.org/10.1016/j.jbankfin.2017.07.002

Westergaard, D., Stærfeldt, H.-H., Tønsberg, C., Jensen, L. J., dan Brunak, S. (2018). A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. *PLOS Computational Biology*, *14*(2), e1005962. https://doi.org/10.1371/journal.pcbi.1005962