

PREDIKSI HARGA SAHAM SEKTOR PERBANKAN LQ45 MENGGUNAKAN BILSTM DENGAN PERBANDINGAN METODE FEATURE SELECTION

Jonathan Wijaya, Muhammad Rizky Pribadi

Informatika, Universitas Multi Data Palembang

Jalan Rajawali No.14 Palembang Indonesia

Jonathanwijaya_2226250009@mhs.mdp.ac.id

ABSTRAK

Pergerakan harga saham yang fluktuatif menjadikan prediksi harga saham sebagai tantangan penting dalam pengambilan keputusan investasi, khususnya pada saham sektor perbankan yang tergabung dalam indeks LQ45. Permasalahan utama dalam prediksi harga saham adalah tingginya volatilitas data serta pemilihan fitur yang kurang optimal sehingga dapat menurunkan akurasi model prediksi. Penelitian ini bertujuan untuk membangun model prediksi harga saham sektor perbankan LQ45 menggunakan algoritma *Bidirectional Long Short-Term Memory* (BiLSTM) serta menganalisis pengaruh penerapan metode seleksi fitur terhadap kinerja model. Metode yang digunakan meliputi penerapan tiga teknik seleksi fitur, yaitu *Spearman Correlation*, *Mutual Information*, dan *Lasso Regression*, dengan dataset historis saham BBRI, BBNI, BMRI, dan BBTN periode 2020–2025. Data dibagi menjadi 80% data pelatihan dan 20% data pengujian, serta dievaluasi menggunakan *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). Hasil penelitian menunjukkan bahwa penerapan seleksi fitur mampu meningkatkan akurasi model BiLSTM, Namun, pada saham BBTN, penerapan *Spearman*, *Mutual information* 10 fitur dan 15 fitur justru menurunkan performa model BiLSTM.

Kata kunci : BiLSTM, Feature Selection, LQ45, Perbankan, Prediksi Harga Saham

1. PENDAHULUAN

Investasi adalah kegiatan menanamkan sejumlah aset dalam suatu instrumen dengan harapan memperoleh keuntungan atau peningkatan nilai di masa depan. Investasi dibagi menjadi dua jenis, yaitu *real investment* yang berupa tanah dan emas, sedangkan *financial investment* berupa saham dan obligasi [1]. Investasi telah menjadi salah satu instrumen penting dalam pengelolaan keuangan baik secara individu maupun dalam dunia bisnis. Selain itu, investasi juga berperan sebagai sarana untuk mencapai tujuan jangka panjang dan terhindar dari inflasi.

Salah satu instrumen yang populer saat ini adalah saham. Saham merupakan instrumen investasi yang bersifat kepemilikan. Saham juga dapat diartikan sebagai bukti pernyataan dalam kepemilikan perusahaan. [2]. Saham menawarkan *return* yang lebih menarik dibandingkan dengan deposito ataupun obligasi. Namun, saham memiliki risiko yang lebih tinggi karena fluktuasi harga yang cukup tinggi pula.

Dalam Bursa Efek Indonesia (BEI) terdapat beberapa indeks seperti IHSG, LQ45, dan IDX30. Indeks LQ45 merupakan indeks yang menghimpun saham – saham paling likuid dan berkapitalisasi besar di Bursa Efek Indonesia. Indeks ini mencerminkan dinamika harga saham unggulan dan direvisi secara berkala berdasarkan kriteria performa, likuiditas, dan kapitalisasi pasar [4]. Berdasarkan karakteristik tersebut, LQ45 menjadi acuan utama bagi investor dalam memantau kinerja saham unggulan. [5]. Terlepas dari banyaknya sisi positif pada indeks LQ45, harga saham dalam indeks ini masih bersifat fluktuatif.

Permasalahan utama dalam memprediksi harga saham sektor perbankan adalah volatilitas tinggi serta

menunjukkan pola momentum dan *reversal*, model BiLSTM dipilih karena mampu mempelajari pola dua arah (*forward* dan *backward*) [7] yang relevan dan adaptif terhadap perubahan pola yang tidak konsisten pada pergerakan harga saham. Model lain seperti LSTM, GRU, ARIMA hanya mampu membaca pola ke depan (*forward*), sehingga tidak digunakan karena tidak mampu menangkap pola sekuensial dan *reversal* secara efektif. Model BiLSTM masih sensitif terhadap fitur tidak relevan, *multicollinearity*, *redundancy* dan *noise* pada data input. Masalah – masalah ini dapat diselesaikan dengan menerapkan *feature selection*.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun sebuah model yang mampu memprediksi harga saham sektor perbankan pada indeks LQ45 menggunakan algoritma BiLSTM dan membandingkan kinerja dari tiap *feature selection* yang diterapkan yaitu *Mutual Information*, *Spearman Correlation* dan *Lasso Regression*. *Feature selection* digunakan untuk mencari fitur paling optimal dan dapat meningkatkan akurasi dalam memprediksi harga saham.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian terdahulu menunjukkan bahwa berbagai algoritma *machine learning* telah digunakan dalam memprediksi harga saham dengan hasil yang beragam. Seperti pada penelitian yang dilakukan oleh Puteri yang membandingkan algoritma LSTM dan BiLSTM dalam memprediksi harga saham. LSTM dan BiLSTM mampu memprediksi harga saham dengan akurasi yang tinggi dengan nilai MAPE < 3%. Namun, BiLSTM mampu memprediksi harga saham dengan

penurunan error nilai MAPE yang lebih kecil pada semua emiten yang diuji [8].

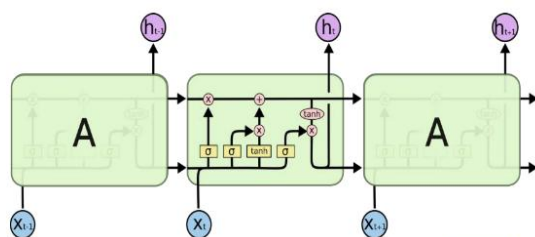
Pada penelitian yang disusun oleh Talib dkk membahas tentang penerapan seleksi fitur pada algoritma *Random Forest* dalam memprediksi harga saham ANTM. Nilai RMSE pada algoritma *Random Forest* adalah 56,629 dan penggunaan seleksi fitur dengan metode *Spearman Correlation* menghasilkan nilai RMSE sebesar 51,820. Penelitian ini membuktikan bahwa penerapan seleksi fitur dapat meningkatkan akurasi prediksi [9]. Studi yang dilakukan oleh Susanto dkk membahas tentang penerapan *Linear Regresion* menggunakan *feature selection* yaitu *mutual information*. Hasil dari penelitian ini adalah nilai *R square* sebesar 0,99455. Nilai ini hampir mendekati 1 yang menandakan bahwa hasil prediksi yang dilakukan hampir mendekati sempurna [10]. Riset yang dilakukan oleh Fitria & Rozci membandingkan antara regresi linear dan regresi lasso. Regresi linear terbukti lebih bagus dengan nilai *R square* sebesar 0,63 dibanding regresi lasso dengan nilai *R square* sebesar 0,62 [11].

2.2. Saham

Saham adalah surat bukti kepemilikan atas aset-aset perusahaan yang menerbitkan saham. Dengan memiliki saham suatu perusahaan maka investor akan mempunyai hak terhadap pendapatan dan kekayaan perusahaan, setelah dikurangi dengan pembayaran semua kewajiban perusahaan [12].

2.3. Long Short Term Memory (LSTM)

LSTM merupakan salah satu pengembangan dari arsitektur *Recurrent Neural Network* (RNN) dengan memberi *memory cell* untuk menyimpan informasi dalam waktu yang lama dan dapat mengatasi terjadinya *vanishing gradient* pada RNN saat melakukan proses data sekuensial yang panjang [13]. Arsitektur LSTM dapat dilihat pada gambar 1.



Gambar 1. Arsitektur LSTM

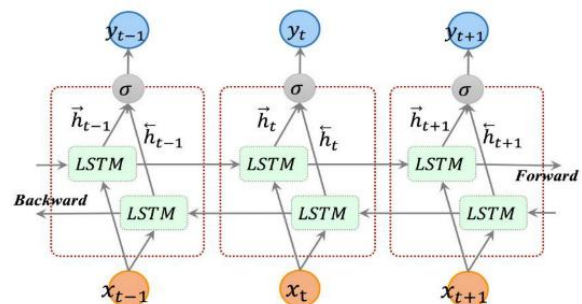
Pada arsitektur LSTM terdapat tiga *gates* yaitu *Forget gate* (f_t), *input gate* (i_t) dan *output gate* (o_t). *Forget gate* berfungsi untuk menghapus informasi dari *cell state*. *Input gate* berfungsi dalam memutuskan informasi baru yang masuk kedalam *cell state*, sedangkan *Output gate* berfungsi dalam memilih informasi yang berguna dari arus *cell state* dan menampilkannya sebagai *output*.

2.4. Bidirectional Long Short Term Memory (BiLSTM)

BiLSTM adalah salah satu penemuan *neural network* yang dapat digunakan untuk pemodelan seri waktu. Selain itu, dapat digambarkan sebagai jaringan saraf karena arsitekturnya yang fleksibel, yang memungkinkan bentuknya berubah sesuai dengan aplikasi yang dibuat [14].

Pada BiLSTM terdapat dua jaringan LSTM dimana jaringan LSTM pertama berfungsi dalam memproses urutan masukan data ke arah depan (*forward*) dan jaringan LSTM kedua berfungsi dalam memproses urutan data dari arah sebaliknya (*backward*). *Output* dari jaringan LSTM *forward* dan *backward* digabungkan pada setiap urutan waktu. Dengan adanya dua *layer* yang berlawanan arah tersebut, model dapat mempelajari informasi masa lalu dan informasi masa mendatang untuk setiap *sequence input* [7]. Berikut adalah persamaan dari BiLSTM (1)

$$h_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (1)$$



Gambar 2. Arsitektur BiLSTM

2.5. Spearman Correlation

Spearman Correlation digunakan untuk mengidentifikasi dan menguji kekuatan suatu hubungan antara dua variabel secara monotonik, baik linier maupun non-linear. Metode ini diharapkan dapat membantu mengidentifikasi fitur-fitur yang paling relevan terhadap variabel target sambil mengurangi kompleksitas model dan potensi *overfitting* [15]. Rumus dari *Spearman Correlation* dapat dilihat pada persamaan (2).

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2)$$

Ket =

ρ = nilai korelasi *Rank Spearman*

d = beda urutan dari setiap pasangan

n = jumlah pasangan

Table 1. Interval Korelasi Rank Spearman

Nilai Interval	Interpretasi
0.90 – 1.00	Korelasi Sangat Kuat
0.70 – 0.89	Korelasi Kuat
0.40 – 0.69	Korelasi Sedang
0.10 – 0.39	Korelasi Lemah
0.00 – 0.09	Korelasi Sangat Lemah

2.6. Mutual Information

Mutual Information (MI) adalah metode untuk menilai seberapa besar informasi yang dimiliki oleh suatu fitur terhadap variable target. Secara umum, metode ini digunakan untuk memahami sejauh mana setiap fitur dalam dataset memberikan informasi yang berguna dalam memprediksi variabel target [10]. Rumus dari Mutual Information dapat dilihat pada persamaan (3).

$$I(X;Y) = \left(\sum_{x \in X} \sum_{y \in Y} p(x,y) \cdot \log \frac{p(x,y)}{p(x) \cdot p(y)} \right) \quad (3)$$

Ket =

$I(X;Y)$ = *mutual information* antara variabel acak X dan Y

$p(x,y)$ = probabilitas hubungan dari X dan Y

$p(x)$ dan $p(y)$ = fungsi probabilitas dari X dan Y

2.7. Lasso Regression

LASSO merupakan sebuah model berbasis supervised regression yang melakukan regularisasi terhadap fitur dan mengidentifikasi fitur yang paling informatif dan paling tidak redundan untuk memprediksi variabel target. Secara matematis, LASSO melakukan penalti dengan cara meminimisasi kuadrat terkecil dengan menggunakan regulasi L1 terhadap koefisien dari variabel prediktor [16]. Penalti ini mendorong model untuk mengurangi koefisien dari fitur yang kurang penting hingga menjadi nol, sehingga secara efektif melakukan seleksi fitur [17]. Rumus dari LASSO regression dapat dilihat pada persamaan (4).

$$\min_{\beta_0, \beta} \left(\frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right) \quad (4)$$

Ket =

N = jumlah observasi atau pengamatan

y_i = respon pada pengamatan ke-i

x_i = sebuah vektor *p-value* sesuai dengan satu nilai Lambda.

λ = parameter regulasi positif yang sesuai dengan satu nilai Lambda

β_0 dan β = Masing – masing merupakan *scaler* dan *p-vector*

2.8. Min – Max Scaler

Min-Max adalah jenis normalisasi yang mentransformasi linier pada data asli untuk menghasilkan data keseimbangan nilai komparatif antara data sebelumnya dan setelah processing [18]. Skala yang dihasilkan berada di antara 0 hingga 1 [19]. Rumus normalisasi Min-Max ditunjukkan pada persamaan (5).

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (5)$$

Ket =

x_i = data yang dinormalisasi

x' = hasil normalisasi

$\min(x)$ = data terkecil suatu fitur

$\max(x)$ = data terbesar suatu fitur

2.9. Optimizer Adam

Adam *optimizer* adalah algoritma optimasi yang sering digunakan dalam deep learning. Adam optimizer memperbarui bobot dan learning rate dapat berubah – ubah selama proses training. Adam optimizer menggabungkan kelebihan dari momentum dan RMSProp *optimizer* yang dapat memecahkan masalah gradien renggang untuk mencapai kinerja yang baik [20]. Adam *optimizer* menggunakan rata – rata momen gradien pertama dan kedua seperti pada RMSProp. Dengan demikian metode ini menggabungkan manfaat dari RMSProp dan momentum [21]. Rumus dari adam dapat dilihat pada persamaan (6) (7) (8) (9)

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (6)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (7)$$

$$m_t^{\hat{}} = \left(\frac{m_t}{1 - \beta_1^t} \right) v_t^{\hat{}} = \left(\frac{v_t}{1 - \beta_2^t} \right) \quad (8)$$

$$\theta_{t+1} = \theta_t - \frac{n}{\sqrt{v_t^{\hat{}} + \epsilon}} m_t^{\hat{}} \quad (9)$$

Ket =

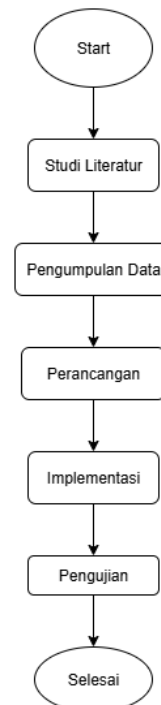
β_1 = laju peluruhan eksponensial dengan nilai default 0.9

β_2 = laju peluruhan eksponensial dengan nilai default 0.999

$m_t^{\hat{}}$ dan $v_t^{\hat{}}$ = bias koreksi untuk m_t dan v_t

3. METODE PENELITIAN

Dalam pelaksanaan penelitian ini, diterapkan perancangan metode terlebih dahulu untuk meminimalisir kesalahan dalam proses penelitian. Tahapan metodologi penelitian dapat dilihat pada Gambar 3.



Gambar 3. Tahapan metodologi penelitian

3.1. Studi Literatur

Tahapan ini dimulai dengan melakukan pembelajaran literatur berupa jurnal sebagai bahan pembelajaran untuk memahami konsep dan teori mengenai penggunaan Bi-directional Long Short Term Memory (BiLSTM) dengan berbagai metode feature selection dalam memprediksi harga saham sektor perbankan pada indeks LQ45. Melalui tahap ini, diperoleh pemahaman mengenai kelebihan dan kekurangan dari setiap metode sebelumnya. Selain itu, studi literatur ini juga membantu dalam mengidentifikasi research gap yang menjadi dasar pengembangan penelitian ini.

3.2. Pengumpulan Data

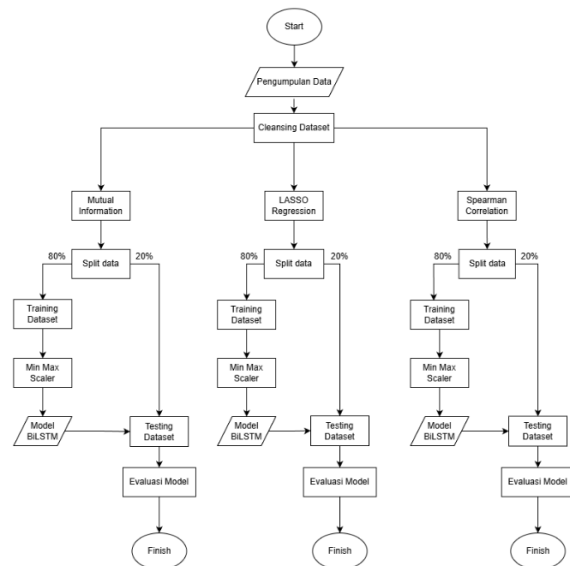
Pada tahap ini, dilakukan proses pengumpulan data historis harga saham dari sektor perbankan yang tergabung dalam indeks LQ45, yaitu BBRI, BBNI, BMRI dan BBTN. Dataset diperoleh dari github yang dikelola oleh wildangunawan dan terdiri dari 1355 data harian periode 29 Juli 2019 hingga 21 Februari 2025 untuk masing – masing emiten. Berikut adalah link dari dataset yang dikelola oleh wildangunawan. <https://github.com/wildangunawan/Dataset-Saham-IDX/tree/master/Saham/LQ45/>

Namun, untuk menjaga konsistensi periode observasi dan membentuk rentang data yang bulat selama lima tahun, maka data yang diambil hanya dari tanggal 21 Februari 2020 hingga 21 Februari 2025 dengan 1210 data harian. Dataset ini memuat 25 atribut yang mencakup informasi harga dan aktivitas perdagangan. Close merupakan variabel target dalam penelitian ini.

3.3. Perancangan

Pada tahapan ini dilakukan perancangan sistem yang dibutuhkan. Diawali dengan mengumpulkan data historis tiap emiten yaitu BBRI, BMRI, BBNI dan BBTN. Setelah data terkumpul, dilakukan *preprocessing data* yang bertujuan untuk menghilangkan data yang kosong dan duplikat. Selanjutnya, dilakukan proses *feature selection* untuk memilih fitur yang paling relevan dan berpengaruh dalam memprediksi harga *close*. Terdapat 3 *feature selection* yang akan digunakan, yaitu *Mutual Information*, *LASSO Regression* dan *Spearman Correlation*. Data dari hasil *feature selection* kemudian dibagi dengan rasio 80% sebagai *data training* dan 20% sebagai *data testing*. 80% *data training* akan dilakukan normalisasi dengan *min max scaler* menjadi skala [0,1]. Hal ini bertujuan untuk mempermudah proses pelatihan model BiLSTM. Model BiLSTM menggunakan fungsi sigmoid dan tanh yang melakukan perhitungan data dengan skala [0,1]. *Data training* akan dilatih menggunakan model BiLSTM. *Optimizer* yang akan digunakan pada model BiLSTM adalah adam. Parameter yang digunakan dalam model BiLSTM adalah *epochs* 200, *batch size* 5, *neuron* 20 dan *timestep* 5. Hasil pelatihan model kemudian akan diuji menggunakan *data testing* untuk

mengevaluasi kinerja model berdasarkan metrik evaluasi, yaitu *Root Mean Square Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). Skema perancangan dapat dilihat pada Gambar 4.



Gambar 4. Skema perancangan model BiLSTM

3.4. Implementasi

Pada tahapan ini dilakukan implementasi dari sistem yang telah dirancang sebelumnya. Implementasi dilakukan pada Google Colab dengan bahasa pemrograman Python. agar sistem dapat mengenali dan melakukan prediksi terhadap data *testing* yang telah dipisahkan sebelumnya. Tahapan implementasi ini memastikan bahwa sistem dapat berjalan sesuai desain dan mampu menghasilkan prediksi berdasarkan data historis yang telah disiapkan.

3.5. Pengujian

Pada tahapan ini, model yang telah dibuat sebelumnya akan melakukan uji coba terhadap data uji. Setelah pengujian berhasil, hasil pengujian akan dievaluasi menggunakan metrik error regresi, yaitu *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage* (MAPE).

RMSE adalah perhitungan akar dari kuadrat error (data aktual – data prediksi) yang dibagi dengan jumlah data (Afrianto et al., 2022). Perhitungan RMSE dapat dilihat pada persamaan (10).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (10)$$

Ket =

y_i = data awal

$\hat{y}_i$ = data akhir

n = jumlah data

MAPE merupakan nilai presentase kesalahan/error yang didapatkan berdasarkan nilai

aktual dengan nilai hasil prediksi (Yang & Wang, 2021) dengan menggunakan persamaan (11).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (11)$$

Ket =

y_i = nilai aktual data

$\hat{y}_i$ = nilai hasil prediksi

n = banyaknya data

4. HASIL DAN PEMBAHASAN

4.1. Data Collection

Dataset diperoleh dari github yang dikelola oleh wildangunawan dan terdiri dari 1210 data harian periode 21 Februari 2020 hingga 21 Februari 2025 untuk masing – masing emiten. Dataset ini memuat 25 atribut yang mencakup informasi harga dan aktivitas perdagangan. Atribut – atribut tersebut adalah date, previous, open_price, first_trade, high, low, close, change, volume, value, frequency, index_individual, offer, offer_volume, bid, bid_volume, listed_shares, tradeable_shares, weight_for_index, foreign_sell, foreign_buy, delisting_date, non_regular_volume, non_regulare_value, non_regular_freuquency. Close merupakan variabel target dalam penelitian ini.

4.2. Pre-Processing Data

Untuk memastikan kualitas data sebelum dilakukan *split data* dan dimasukkan ke dalam model, maka dilakukanlah *pre-processing data*. Pada tahap ini dilakukan pengecekan nilai kosong dan duplikat pada dataset. Selain itu karena penelitian ini berbasis *time series*, maka dilakukan pengecekan urutan dari tiap baris data.

4.3. Spearman Correlation

Spearman digunakan untuk mencari fitur yang memiliki kekuatan hubungan yang kuat secara *monotonic*. Fitur yang akan digunakan pada model adalah fitur yang memiliki nilai > 0.90 yang memiliki korelasi sangat kuat.

4.4. Mutual Information

Mutual Information digunakan untuk mencari fitur yang relevan dan mengevaluasi jumlah informasi yang dipelajari tentang variable target dari tiap fitur

yang ada, Pemilihan fitur akan dibagi menjadi 3 skenario, yaitu 5 fitur dengan skor teratas, 10 fitur dengan skor teratas dan 15 fitur dengan skor teratas. 3 skenario ini akan dibandingkan dan dapat menjadi gambaran seberapa banyak jumlah fitur yang dapat meningkatkan akurasi

4.5. LASSO Regression

LASSO digunakan untuk mengidentifikasi fitur yang paling informatif dan paling tidak redundan untuk memprediksi variabel target. LASSO akan melakukan penalti dan mendorong model untuk mengurangi koefisien dari fitur yang kurang penting hingga menjadi nol. Maka, fitur yang akan digunakan pada model adalah semua fitur yang memiliki nilai. Karena, fitur yang memiliki nilai adalah fitur yang paling informatif

4.6. Bidirectional Long Short Term Memory

BiLSTM digunakan sebagai model utama dalam memprediksi harga penutupan (*close*) pada penelitian ini. BiLSTM mampu bekerja dengan dua arah yaitu *forward* dan *backward*. Pendekatan dua arah ini membuat BiLSTM mampu mempelajari depedensi jangka panjang dan dapat membaca perubahan pola (*reversal*) atau tren yang terjadi pada harga saham.

Dalam penelitian ini, BiLSTM menerima input fitur – fitur terpilih dari proses *feature selection* (*Spearman Correlation*, *Mutual Information* dan *LASSO Regression*). Parameter yang akan digunakan dalam model BiLSTM ini adalah epoch 500, batch size 256, neuron 20 dan timestep 5. Optimizer yang digunakan pada model BiLSTM adalah adam.

4.7. Hasil Evaluasi

Hasil evaluasi model diukur menggunakan dua metrik, yaitu *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). Dalam penelitian ini, MAPE dijadikan indikator utama untuk menilai apakah penerapan teknik *feature selection* dapat meningkatkan akurasi prediksi. Melalui perbandingan nilai MAPE pada setiap skenario, dapat dilihat apakah proses seleksi fitur memberikan manfaat signifikan atau justru menunjukkan bahwa model tanpa *feature selection* sudah bekerja secara optimal.

Tabel 1. Hasil Evaluasi Model BiLSTM dan Feature Selection

Feature Selection		BBRI		BBNI		BMRI		BBTN	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Tanpa Feature Selection		169.62	2.52%	138.73	2.32%	219.78	2.54%	38.05	2.04%
Spearman Correlation		120.72	1.85%	141.06	2.25%	172.3	2.16%	40.44	2.22%
LASSO Regression		129.42	2.10%	137.48	2.11%	144.51	1.81%	36.69	1.92%
Mutual Information	5 fitur	118.95	1.92%	132.83	2.4%	169.24	2.11%	36.03	1.92%
	10 fitur	115.47	1.84%	126.26	2.03%	207.96	2.50%	41.84	2.31%
	15 fitur	122.87	1.96%	142.75	2.26%	168.03	2.04%	44.89	2.59%

Hasil evaluasi model BiLSTM menunjukan bahwa penerapan *feature selection* memberikan pengaruh yang cukup signifikan terhadap akurasi prediksi harga saham pada empat emiten sektor

perbankan LQ45, yaitu BBRI, BBNI, BMRI, dan BBTN. Model tanpa *feature selection* secara konsisten menghasilkan nilai MAPE yang lebih tinggi dibandingkan sebagian besar metode *feature selection*

lainnya, kecuali pada emiten BBTN, dimana nilai MAPE pada *Spearman Correlation*, *Mutual information* 10 fitur dan 15 fitur lebih tinggi dibandingkan tanpa *feature selection*.

Spearman Correlation memberikan peningkatan akurasi yang cukup baik pada emiten BBRI yaitu peningkatan sebesar 0,67% dan pada emiten BBRI dengan peningkatan sebesar 0.38%. *LASSO Regression* menunjukan performa yang stabil dan mampu menurunkan error pada semua emiten dibandingkan model *baseline*. *Mutual Information* dengan 10 fitur merupakan konfigurasi yang paling efektif untuk sebagian besar emiten, khususnya pada BBRI dan BBNI. Penambahan fitur menjadi 15 justru menurunkan performa karena kemungkinan penambahan fitur kurang relevan justru menambah *noise*. Kecuali pada emiten BMRI yang memiliki performa terbaik dengan penggunaan 15 fitur *Mutual Information*.

Hanya emiten BBTN yang mengalami penurunan performa saat menerapkan *feature selection*. *Spearman*, *Mutual information* 10 fitur dan 15 fitur mengalami penurunan, sementara pada *LASSO* dan *Mutual Information* 5 fitur tetap mengalami peningkatan. Dapat disimpulkan bahwa BBTN membutuhkan jumlah fitur yang lebih sedikit atau pemilihan fitur berbasis regularisasi agar model dapat bekerja dengan optimal.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil membangun model prediksi harga saham menggunakan algoritma BiLSTM dan mengevaluasi pengaruh berbagai metode *feature selection* terhadap akurasi prediksi pada emiten sektor perbankan LQ45, yaitu BBRI, BBNI, BMRI, dan BBTN. Hasil penelitian meunjukkan bahwa penerapan *feature selection* secara umum mampu meningkatkan akurasi model dibandingkan penggunaan seluruh fitur secara langsung. *Spearman Correlation* memberikan peningkatan signifikan pada emiten tertentu namun kurang konsisten pada emiten lainnya. *LASSO Regression* memberikan performa stabil dan berhasil meningkatkan akurasi pada semua emiten. *Mutual information* dengan skenario 10 fitur menjadi metode yang paling efektif untuk emiten BBRI dan BBNI. Pengecualian terjadi pada emiten BBTN, di mana beberapa metode *feature selection* yaitu *Spearman Correlation*, *Mutual Information* 10 dan 15 fitur justru menurunkan akurasi.

Secara keseluruhan, penelitian ini menegaskan bahwa *feature selection* memegang peran penting dalam meningkatkan performa BiLSTM dalam memprediksi harga saham. Namun, efektivitas metode *feature selection* masih bergantung pada karakteristik data pada masing – masing emiten.

DAFTAR PUSTAKA

[1] S. E. Destina Paningrum, *Buku referensi investasi pasar modal*. Lembaga Chakra Brahmana Lentera, 2022.

[2] A. L. Putra and A. Kurniawati, “Analisis Prediksi Harga Saham PT. Astra International Tbk Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) dan Support Vector Regression (SVR),” *J. Ilm. Komputasi*, vol. 20, no. 3, pp. 417–423, 2021, doi: 10.32409/jikstik.20.3.2732.

[3] CNN, “Investor Pasar Modal Tembus 14,87 Juta pada 2024, Naik 6 Kali Lipat.” [Online]. Available: <https://www.cnnindonesia.com/ekonomi/20250211123953-92-1196941/investor-pasar-modal-tembus-1487-juta-pada-2024-naik-6-kali-lipat>

[4] S. Wijaya and G. S. S. Ugut, “Tata Kelola dan Performa Finansial Perusahaan di Indeks LQ45,” *J. Maksipreneur Manajemen, Koperasi, dan Entrep.*, vol. 14, no. 1, pp. 97–114, 2024, doi: 10.30588/jmp.v14i1.1612.

[5] L. Ismawati, J. D. Asri, and I. Yuliastuti, “PENETAPAN INDEKS HARGA SAHAM LQ45 DENGAN PERTIMBANGAN TINGKAT INFLASI DAN TINGKAT SUKU BUNGA SERTIFIKAT BANK INDONESIA (SBI) PADA BURSA EFEK INDONESIA (BEI) DETERMINATION OF THE LQ45 STOCK PRICE INDEX WITH CONSIDERATION OF THE INFLATION RATE AND THE B,” vol. 51, no. 1, pp. 98–108, 2025.

[6] S. Kusumawati, “DAMPAK DIUMUMKANNYA KASUS COVID-19 SERTA KEBIJAKAN NEW NORMAL TERHADAP PERUBAHAN HARGA DAN VOLUME,” vol. 19, 2020.

[7] Y. Karyadi, “Prediksi Kualitas Udara Dengan Metoda LSTM, Bidirectional LSTM, dan GRU,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 671–684, 2022, doi: 10.35957/jatisi.v9i1.1588.

[8] D. I. Puteri, “Implementasi Long Short Term Memory (LSTM) dan Bidirectional Long Short Term Memory (BiLSTM) Dalam Prediksi Harga Saham Syariah,” *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 11, no. 1, pp. 35–43, 2023, doi: 10.34312/euler.v11i1.19791.

[9] E. A. H. Talib, A. F. Watratan, and Saharuddin, “COMPARATIVE ANALYSIS OF THE APPLICATION OF FEATURE SELECTION IN RANDOM FOREST REGRESSION FOR STOCK PRICE PREDICTION,” *Nusant. Hasana J.*, vol. 5, no. 3, pp. 334–348, 2025, [Online]. Available: <https://www.nusantarahasanaajournal.com/index.php/nhj/article/view/1641/1441>

[10] A. L. Susanto, N. W. S. Saraswati, M. W. Adhiputra, and I. D. M. K. Muku, “Prediksi Harga Bitcoin Menggunakan Metode Linear Regresion,” *J. Math. Theory Appl.*, vol. 13, no. 2, pp. 43–52, 2024, doi: 10.32938/j-math.v2i2.7699.

[11] E. R. Fitria and F. Rozci, “Penerapan Metode

- Regresi Least Absolute Shrinkage and Selection Operator (LASSO) dan Regresi Linier untuk Memprediksi Tingkat Kemiskinan di Indonesia,” *J. Ilm. Sosio Agribis*, vol. 22, no. 2, p. 123, Dec. 2023, doi: 10.30742/jisa22220222620.
- [12] A. A. Kurniasi, M. A. Saptari, and V. Ilhadi, “Aplikasi Peramalan Harga Saham Perusahaan Lq45 Dengan Menggunakan Metode Arima,” *Sisfo J. Ilm. Sist. Inf.*, vol. 5, no. 1, pp. 13–26, 2021, doi: 10.29103/sisfo.v5i1.4849.
- [13] R. S. Pontoh *et al.*, “Jakarta Pandemic to Endemic Transition: Forecasting COVID-19 Using NNAR and LSTM,” *Appl. Sci.*, vol. 12, no. 12, 2022, doi: 10.3390/app12125771.
- [14] E. Subowo, F. Adi Artanto, I. Putri, and W. Umaedi, “Algoritma Bidirectional Long Short Term Memory untuk Analisis Sentimen Berbasis Aspek pada Aplikasi Belanja Online dengan Cicilan,” *J. Fasilkom*, vol. 12, no. 2, pp. 132–140, 2022, doi: 10.37859/jf.v12i2.3759.
- [15] J. Bocianowski, D. Wrońska-Pilarek, A. Krysztofiak-Kaniewska, K. Matusiak, and B. Wiatrowska, “Comparison of Pearson’s and Spearman’s correlation coefficients for selected traits of *Pinus sylvestris* L.,” *Biometrical Lett.*, vol. 61, no. 2, pp. 115–135, 2024, doi: 10.2478/bile-2024-0008.
- [16] T. J. Nigon, C. Yang, G. D. Paiao, D. J. Mulla, J. F. Knight, and F. G. Fernández, “(利用高分辨率航空高光谱影像预测玉米早季氮素吸收),” *Remote Sens.*, vol. 12, no. 8, 2020.
- [17] H. Dineep, M. Kamalesh, A. Prasanna T, A. KK, S. K, and J. Kannan R, “Household Energy Bill Prediction Using Various Machine Learning Techniques,” *SSRN Electron. J.*, 2025, doi: 10.2139/ssrn.5183320.
- [18] R. Chandra, K. Chaudhary, and A. Kumar, “Comparison of Data Normalization for Wine Classification Using K-NN Algorithm,” vol. 5, no. 4, pp. 175–180, 2022.
- [19] P. P. Allorerung, A. Erna, and M. Bagussahrir, “Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit,” vol. 9, no. 3, pp. 178–191, 2024.
- [20] F. Ramadhan and J. Hernadi, “Evaluasi optimizer adam dan rmsprop pada arsitektur vgg-19 klasifikasi ekspresi wajah manusia,” vol. 10, no. 2, pp. 1414–1426, 2025.
- [21] M. Reyad and A. M. Sarhan, “A modified Adam algorithm for deep neural network optimization,” *Neural Comput. Appl.*, vol. 35, no. 23, pp. 17095–17112, 2023, doi: 10.1007/s00521-023-08568-z.