

Optimizing Indonesian Banking Stock Predictions with DBSCAN and LSTM

Septiannisa Alya Shinta Purwandhani¹, Aletta Agigia Novta Sajiatmoko², Christian Sri Kusuma Aditya^{*3}

^{1,2,3}Informatics, Universitas Muhammadiyah Malang, Indonesia

Email: christianskaditya@umm.ac.id

Received : Feb 19, 2025; Revised : Apr 25, 2025; Accepted : May 12, 2025; Published : Jun 10, 2025

Abstract

Investing in the stock market is challenged by high volatility, which often leads to inaccurate price predictions. Prediction models often struggle to handle the fluctuation phenomenon and produce unstable forecasts. This study aims to predict stock prices in three banks, namely PT Bank Central Asia Tbk (BBCA), PT Bank Rakyat Indonesia (Persero) Tbk (BBRI), and PT Bank Mandiri (Persero) Tbk (BMRI) using Long Short-Term Memory (LSTM) with the integration of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for anomaly detection. DBSCAN is applied with an epsilon (ϵ) of 0.5 and a minimum of 5 samples using Euclidean distance. The LSTM model consists of two hidden layers with 50 units, optimized using Adam, and applying the Mean Squared Error (MSE) loss function. The results show that DBSCAN improves prediction accuracy under several conditions. For BBCA stock, the lowest MSE was 0.003 at the 2nd fold with DBSCAN compared to 0.006 without DBSCAN. For BMRI stock achieved an MSE of 0.003 at the 4th fold with DBSCAN, while the 5th fold without DBSCAN obtained 0.000. For BBRI stock showed the best MSE of 0.003 at the 2nd fold with DBSCAN and the 5th fold without DBSCAN. These results show that the integration of DBSCAN can improve prediction especially when extreme price fluctuations occur. This research contributes to the development of stock price prediction methods that can be one of the benchmarks for investors before making decisions so that they do not experience losses.

Keywords : Anomaly, Banking, DBSCAN, LSTM, Outlier, Stock.

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License

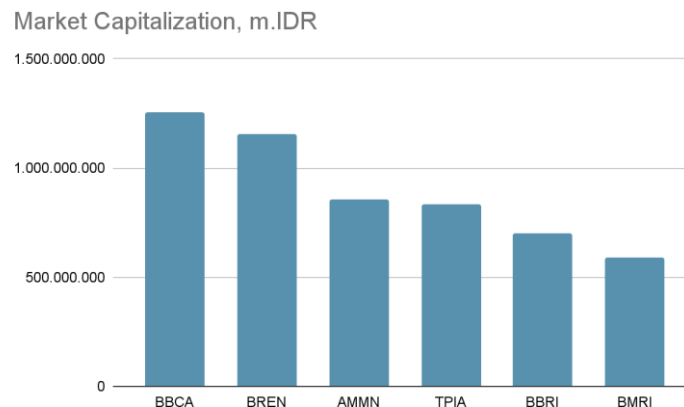


1. PENDAHULUAN

Investasi telah menjadi topik yang banyak diperbincangkan oleh masyarakat umum maupun institusi besar dalam beberapa tahun terakhir. Di era digital saat ini, investasi tidak hanya berfungsi sebagai sarana untuk meningkatkan aset, tetapi juga sebagai alat untuk mencapai tujuan keuangan jangka panjang. Jenis investasi pun beragam, mulai dari investasi dalam bentuk aset nyata (riil assets) seperti emas, barang berharga, tanah, karya seni, atau properti, hingga investasi dalam bentuk surat berharga atau sekuritas (*marketable securities atau financial assets*) [1]. Potensi keuntungan tinggi yang ditawarkan serta kemudahan akses melalui berbagai macam *platform* daring membuat saham menjadi salah satu instrument investasi yang paling diminati [2]. Saham sendiri dapat diartikan sebagai kepemilikan seorang investor dalam suatu perusahaan, tanda investasi yang dimilikinya, atau jumlah modal yang ditanamkan dalam entitas bisnis tertentu [3]. Dalam sektor perbankan, saham memiliki peran penting bagi perekonomian, khususnya di Indonesia. Keuntungan yang diperoleh dari investasi saham dapat berupa *capital gain* serta dividen yang dibagikan kepada pemegang saham dalam jangka waktu tertentu, baik secara berkala maupun tahunan [4]. Menurut William A. Lovett [5], sektor perbankan merupakan inti dari perekonomian suatu negara. Termasuk di Indonesia, di mana perbankan memainkan peranan krusial dalam pertumbuhan ekonomi nasional.

Potensi keuntungan besar yang ditawarkan dari saham juga tidak lepas dari risiko kerugian, instrumen ini juga dikenal memiliki risiko yang tinggi dan fluktuatif. Perubahan harga saham yang terjadi secara cepat dan tidak terduga dapat memberikan pengaruh yang signifikan terhadap keputusan investasi baik oleh individu maupun institusi keuangan [6]. Oleh karena itu, saat akan mengambil keputusan investasi, investor biasanya melakukan analisis terhadap aspek-aspek Perusahaan. Hal tersebut dikarenakan kondisi keuangan suatu perusahaan berkaitan erat dengan pergerakan harga sahamnya, semakin besar pendapatan yang diperoleh perusahaan, semakin tinggi pula nilai sahamnya [7]. Fenomena ini menegaskan pentingnya pengembangan model prediksi yang dapat diandalkan untuk menghadapi ketidakpastian pasar saham, khususnya di Indonesia. Prediksi yang akurat serta deteksi dini terhadap anomali dalam data saham menjadi faktor kunci yang dapat membantu investor dalam membuat keputusan yang lebih tepat.

Menurut Putri & Dhini (2019), Bursa Efek Indonesia (BEI) atau biasa juga dikenal sebagai *Indonesia Stock Exchange* (IDX) merupakan pasar saham utama Indonesia [8]. Dalam penelitian ini, digunakan data dari tiga emiten utama di Indonesia, yaitu PT Bank Central Asia Tbk. (BBCA), PT Bank Rakyat Indonesia (Persero) Tbk. (BBRI), dan PT Bank Mandiri (Persero) Tbk. (BMRI). Ketiga saham ini menempati peringkat teratas dalam daftar 50 saham terbaik di BEI pada Juli 2024 [9]. Pemilihan saham-saham ini didasarkan pada kapitalisasi pasar yang besar serta volume perdagangan yang tinggi, yang mencerminkan performa yang stabil dan peran pentingnya dalam sektor perbankan nasional. Dataset yang digunakan dalam penelitian ini mencakup data historis saham selama beberapa tahun, dengan berbagai atribut seperti harga pembukaan, harga penutupan, volume perdagangan, dan lain-lain. Analisis data ini diharapkan dapat memberikan wawasan yang lebih menyeluruh terkait pergerakan harga saham serta faktor-faktor yang memengaruhinya. Grafik saham dapat dilihat pada Gambar 1.



Gambar 1. Highest Stock Chart as of July 2024 (<https://idx.co.id>)

Dalam analisis data saham, sering kali ditemukan adanya *outlier* yang muncul akibat fluktuasi harga atau volume perdagangan yang ekstrem dalam periode tertentu. *Outlier* yang terdapat pada data harus ditangani. Jika tidak ditangani dengan baik, keberadaan *outlier* pada data dapat menurunkan akurasi model prediksi dan menghasilkan analisis yang kurang akurat. Selain itu, *outlier* juga dapat mempengaruhi proses pengambilan keputusan dan kesimpulan karena menyebabkan distribusi data menjadi tidak normal. Kemunculan *outlier* dapat disebabkan oleh beberapa hal. Faktor eksternal seperti krisis ekonomi, berita keuangan yang signifikan, atau kejadian khusus yang mempengaruhi pasar dapat menjadi penyebab utama munculnya anomali ini. [10]. Untuk itu, pendeteksian dan penanganan anomali dalam data menjadi langkah yang sangat penting untuk meningkatkan kualitas model prediktif. Deteksi anomali merupakan teknik yang banyak digunakan dalam penelitian, karena mampu mengidentifikasi

data yang menyimpang dari pola umum. Teknik ini menjadi bagian esensial dalam pemrosesan data untuk memastikan kualitasnya tetap terjaga [11].

Salah satu teknik yang dapat digunakan untuk mendeteksi anomali pada data adalah *clustering* [12]. *Clustering* merupakan metode data mining yang bersifat *unsupervised* atau tanpa arahan. Cara kerja dari metode ini adalah berdasarkan dengan *similarity* atau kesamaan karakteristik pada data [13]. Data dikelompokkan berdasarkan kepadatan, data dengan kepadatan tinggi akan dikelompokkan secara bersama dan data dengan kepadatan rendah maka dianggap sebagai anomali/*outlier* [14]. Penelitian ini menerapkan DBSCAN (*Density-Based Spatial Clustering of Application with Noise*) dalam mendeteksi *outlier*. DBSCAN mampu mendeteksi *outlier* secara optimal dan memiliki kemampuan fleksibel dalam menangani kepadatan data yang bervariasi, yang membuat DBSCAN digunakan sebagai alat yang berguna untuk menganalisis data kompleks [15]. DBSCAN menerapkan parameter epsilon (jarak antar poin) dan MinPts (minimal poin agar terbentuk *cluster*) untuk mengelompokkan data [16].

Berdasarkan hasil deteksi anomali yang dilakukan pada dataset yang digunakan pada penelitian ini, tidak ditemukan anomali atau *outlier*. Dari ketidakadaan *outlier* tersebut dapat disebut bahwa data yang digunakan stabil dan tidak terdapat perubahan atau lonjakan yang terjadi secara signifikan yang menunjukkan bahwa pada pola data tidak terdapat penyimpangan maupun variasi yang besar dan bergerak secara teratur serta konsisten. Hal ini menandakan bahwa dataset dapat digunakan tanpa perlu melewati penanganan khusus terlebih dahulu terhadap *outlier* atau anomali.

Selain DBSCAN metode *Deep Learning* juga digunakan pada penelitian ini. Faktor-faktor seperti data historis, volume perdagangan, sentimen pasar, dan indikator lainnya membuat dinamika pasar menjadi kompleks [17]. LSTM (*Long Short-Term Memory*) dapat digunakan untuk mengatasi kompleksitas tersebut. LSTM merupakan turunan dari RNN (*Recurrent Neural Network*) dengan memodifikasinya dengan menambahkan *memory cell* yang berguna untuk menyimpan informasi dalam jangka panjang [18]. Kemampuan dalam mengingat informasi dan mengolah permasalahan pada jangka waktu yang panjang pada data *time series* dan pemanfaatan teknologi *Artificial Intelligence* (AI) dan *machine learning* metode ini dapat digunakan untuk memprediksi pergerakan harga saham [19]–[21].

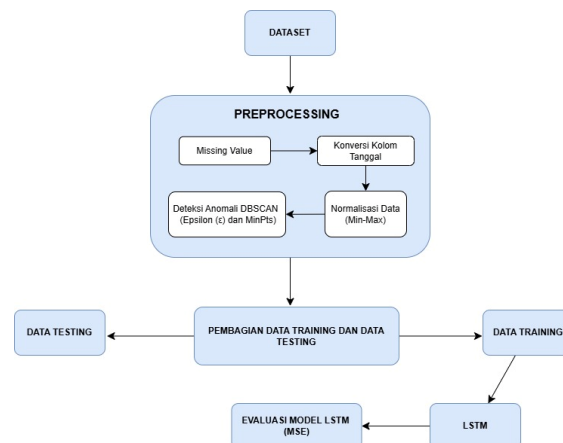
Beberapa penelitian sebelumnya telah melakukan deteksi anomali menggunakan metode seperti *Local Outlier Factor* (LOF), DBSCAN, dan *Isolation Forest*. Sebuah studi menerapkan metode DBSCAN dan telah didapatkan dan dicatat bahwa performa terdeteksi *outlier* adalah sebesar 5% [22]. Namun, penelitian tersebut juga mencatat bahwa deteksi anomali dengan metode saat ini belum dapat menangkap informasi implisit yang memerlukan interpretasi ahli domain. Selain itu, penelitian mengenai prediksi harga saham dengan model LSTM telah menunjukkan keunggulannya dalam menangani pola data non-linear serta tren tahunan. LSTM, yang merupakan pengembangan dari arsitektur RNN, memiliki kemampuan yang baik dalam mengolah data yang bersifat multi-periode [23]. Penelitian sebelumnya menunjukkan bahwa model LSTM dalam memprediksi harga saham BBRI menghasilkan nilai RMSE sebesar 227,47 [24]. Sementara itu, studi lain mengenai optimasi *deep learning* untuk prediksi saham selama pandemi Covid-19 melaporkan hasil RMSE sebagai berikut: BBKA sebesar 1120,65; BBRI sebesar 156,29; ASII sebesar 134,55; dan TLKM sebesar 71,65 [25].

Penelitian ini berfokus pada pengembangan model prediksi harga saham dengan menggunakan LSTM serta pengoptimalan metode DBSCAN untuk deteksi anomali dalam dataset saham BBKA, BBRI, dan BMRI. Optimalisasi DBSCAN dilakukan dengan menerapkan pendekatan *Nearest Neighbors*. Tidak hanya mampu mendeteksi anomali dan menanganinya, penelitian ini juga mampu mengidentifikasi pola dan tren pada harga saham perbankan dengan akurasi yang baik, sehingga dapat mendapatkan hasil analisis yang lebih menyeluruh. Tujuan utama penelitian ini adalah memberikan wawasan bagi investor dalam mengenali pola harga saham serta mendeteksi potensi anomali yang dapat mempengaruhi keputusan investasi. Hasil dari penelitian ini diharapkan mampu meningkatkan akurasi

prediksi harga saham serta memperkaya pemahaman mengenai penerapan *machine learning* dalam analisis data keuangan di Indonesia.

2. METODE PENELITIAN

Pada penelitian kali ini terbagi menjadi lima yaitu pengambilan data, *preprocessing data*, deteksi anomali menggunakan DBSCAN, modeling dengan menggunakan LSTM, dan analisis hasil. Alur untuk penelitian ini dapat dilihat pada Gambar 2.



Gambar 2. Research Flow

2.1. Pengambilan Data

Penelitian ini menggunakan data yang diperoleh dari situs resmi Kaggle, yang berisi informasi mengenai harga saham harian. Data tersebut terdiri dari 10 dataset terpisah, namun penelitian ini hanya berfokus pada 3 dataset utama berdasarkan peringkat 50 besar saham di situs resmi IDX pada Juli 2024 [26]. Setiap dataset merepresentasikan 10 saham paling populer di pasar saham Indonesia. Masing-masing dataset mencakup sekitar 4.000 catatan transaksi harian per saham, mencakup periode historis dari 14 Juli 2003 hingga 3 Juli 2020. Informasi yang tersedia dalam dataset ini meliputi tanggal transaksi saham (*Date*), harga saat pembukaan (*Open*), harga tertinggi sepanjang hari (*High*), harga terendah (*Low*), harga penutupan (*Close*), harga penutupan yang telah disesuaikan untuk faktor-faktor seperti dividen dan pemecahan saham (*Adj Close*), serta volume transaksi (*Volume*). Seluruh data kemudian disimpan dalam format *Comma Separated Values* (CSV) dan diarsipkan dalam *drive* yang telah disiapkan oleh peneliti. Berikut beberapa sampel data yang akan digunakan dalam penelitian ini dapat dilihat pada Tabel 1, Tabel 2, dan Tabel 3.

Tabel 1. BBRI Data Sample

Date	Open	High	Low	Close	Adj Close	Volume
2003-07-14	381.000	430.100	381.000	417.900	134.300	1198338043
2003-07-15	430.100	442.400	417.900	417.900	134.300	420912767
2003-07-17	405.600	417.900	405.600	417.900	134.300	111095912
2003-07-18	405.600	417.900	405.600	417.900	134.300	41139892

Tabel 2. BMRI Data Sample

Date	Open	High	Low	Close	Adj Close	Volume
2003-11-10	105.000	110.000	95.000	97.500	1.373	5.144140e+09
2003-11-11	97.500	100.000	95.000	100.000	1.408	2.938555e+09
2003-11-12	97.500	105.000	97.5000	105.000	1.478	2.002915e+09
2003-11-13	105.000	105.000	102.500	105.000	1.478	1.190050e+09

Tabel 3. BBCA Data Sample

Date	Open	High	Low	Close	Adj Close	Volume
2004-06-08	875.000	900.000	875.000	887.500	438.253	99830000.0
2004-06-09	887.500	912.500	875.000	900.000	444.425	58858000.0
2004-06-10	900.000	900.000	887.500	900.000	444.425	33118000.0
2004-06-11	887.500	900.000	887.500	887.500	444.425	27166000.0

Fitur harga tutup (*'Close'*) merupakan fitur yang dipilih dari beberapa fitur yang ada pada dataset dan akan dijadikan sebagai target data pada penelitian ini.

2.2. Preprocessing Data

Terdapat beberapa Langkah yang dilakukan pada *step preprocessing* ini. Langkah pertama yang dilakukan yaitu *cleaning data*. *Data cleaning* atau pembersihan data merupakan proses yang digunakan untuk mendeteksi, memperbaiki dan juga menangani dataset, tabel, serta database yang tidak lengkap atau tidak akurat. Konsep ini mengacu pada *dirty data* yang akan diganti, dimodifikasi atau dihapus setelah tahap identifikasi data yang tidak menyeluruh, tidak valid, tidak sesuai, dan tidak relevan [27]. Pada penelitian ini cek *missing values* adalah langkah awal yang dilakukan pada *preprocessing data*. *Missing Values* merupakan suatu permasalahan dimana terdapat beberapa bagian data yang hilang [28]. untuk mengidentifikasi apakah ada nilai yang hilang tersebut, seperti menghapus baris atau kolom yang memiliki *missing values*, atau mengisinya dengan nilai tertentu (imputasi).

Setelah menangani *missing values*, langkah selanjutnya adalah melakukan *transformation* data yang meliputi normalisasi data menggunakan *min-max scaling* dan mengonversikan kolom *'Date'* ke format *'DateTime'*. Duplikasi pada data sering terjadi, karena itu terdapat metode yang dapat digunakan untuk menghindari hal tersebut yang disebut metode normalisasi. Normalisasi adalah parameter digunakan untuk mencegah duplikasi terhadap tabel dalam data. Selain itu, normalisasi juga merupakan proses mengklasifikasikan sebuah tabel yang masih memiliki beberapa anomali atau ketidak wajarannya sehingga menghasilkan tabel yang lebih sederhana dan struktur yang bagus [29]. Proses normalisasi data ini sangat diperlukan untuk membentuk data dalam kondisi nilai dengan rentang yang sama, agar dapat menghilangkan data *outlier* dan membuat distribusi data menjadi normal. Pada penelitian ini tahap normalisasi yang dilakukan adalah dengan mengimplementasikan *min-max scaling*. *Min-max scaling* adalah salah satu teknik normalisasi yang berfungsi untuk mentransformasikan data secara linear dengan memakai nilai minimal dan nilai maksimal pada data yang ada. Kemudian dari proses transformasi itu akan dihasilkan keseimbangan antar data yang ada pada kisaran rentang nilai yang sama [30]. Untuk mendapatkan nilai *min-max scaling* digunakan persamaan (1):

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

Setelah tahap normalisasi selesai dilakukan, langkah selanjutnya adalah mengonversi kolom *'Date'* ke format *'DateTime'* agar data dapat digunakan dalam analisis time series. Tahap akhir dalam preprocessing adalah deteksi anomali menggunakan metode DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) untuk mengidentifikasi data yang dianggap *outliers* dalam dataset yang telah dinormalisasi. Proses ini dimulai dengan membaca dataset dan memilih fitur utama seperti *'Open'*, *'High'*, *'Low'*, *'Close'*, dan *'Volume'* sebagai input untuk model. Dalam penerapan DBSCAN pada penelitian ini, terdapat dua parameter utama yang digunakan, yaitu epsilon (ϵ) dan MinPts (*Minimum Points*). Parameter ϵ menentukan radius maksimum yang digunakan untuk mengevaluasi

kedekatan antar data, sedangkan *MinPts* adalah jumlah minimum titik dalam radius tersebut agar suatu titik dapat dianggap sebagai bagian dari *cluster*. Titik yang tidak memiliki cukup tetangga dalam radius ϵ akan dianggap sebagai *outlier*. Pemilihan nilai ϵ dapat dilakukan menggunakan metode *k-distance graph* atau *elbow method* untuk menentukan nilai optimal. Setelah parameter ditetapkan, model DBSCAN diterapkan untuk mengelompokkan data, di mana titik-titik yang tidak masuk dalam *cluster* utama akan dianggap sebagai anomali. Hasil dari proses ini kemudian divisualisasikan untuk melihat distribusi *cluster* dan data *outliers*, yang biasanya ditandai dengan warna berbeda dari kelompok utama. Dengan menggunakan metode DBSCAN, deteksi anomali dapat dilakukan secara efektif berdasarkan kepadatan data, sehingga membantu dalam mengidentifikasi pergerakan harga saham yang tidak normal.

Selanjutnya, dataset yang telah di-*cluster* dipecah menjadi fitur ('x') dan target ('y'). Kolom yang tidak relevan, seperti 'Date', 'Close', dan 'Cluster', dihapus untuk membentuk dataset fitur. Dataset ini kemudian disimpan dalam file terpisah untuk mempermudah proses selanjutnya. Data yang telah dipisahkan kemudian dibagi menjadi data *training* dan *testing* menggunakan metode *split 70-30* atau *K-Fold Cross Validation* untuk memastikan model mendapatkan evaluasi yang baik. Data hasil *split* ini disimpan dan siap digunakan untuk pelatihan model, yaitu LSTM serta evaluasi kinerjanya.

2.3. Deteksi Anomali DBSCAN

Terdapat satu teknik penting yang sering digunakan pada data mining, yaitu teknik Spatial data *clustering*. Teknik ini digunakan untuk memperoleh wawasan atau pemahaman pada data spasial dalam jumlah yang besar dari berbagai aplikasi. Salah satu teknik yang menjadi pelopor dalam perkembangan algoritma *clustering* pada data spasial adalah *Density-Based Spatial Clustering Algorithm with Noise* atau lebih dikenal dengan sebutan DBSCAN. DBSCAN biasanya diaplikasikan pada data-data spasial objek 2 dimensi [31]. DBSCAN bekerja dengan cara mengelompokkan dan mengidentifikasi *cluster* berdasarkan dengan kepadatan titik data.

Terdapat 2 parameter yang ada pada DBSCAN, yaitu *Eps* dan *MinPts*. *Eps* merupakan radius maksimum pada suatu daerah (*neighborhood*), sedangkan *MinPts* merupakan jumlah minimum titik yang harus terdapat di radius *Eps-neighborhood* dari suatu titik untuk membangun sebuah *cluster*. Terdapat beberapa istilah penting pada konsep dasar *density-based clustering*. *Neighborhood* yang ada pada dalam radius *Eps* disebut dengan *Eps-neighborhood* suatu titik. Jika *Eps-neighborhood* yang telah memenuhi nilai *MinPts* yang di dalamnya terdapat suatu titik yang memiliki jumlah *neighborhood*, maka titik tersebut dinamakan *core object*. Pada Kumpulan objek D, terdapat sebuah titik p yang dimana titik p dapat dianggap sebagai *directly density-reachable* dari titik q jika titik p sudah berada di dalam *Eps-neighborhood* dari q dan q adalah *core object* [32].

Pada DBSCAN terdapat metode perhitungan jarak. Salah satu jenis perhitungan jarak yaitu *Euclidean Distance* yang dimana metode tersebut menghitung jarak antara dua buah titik yang ada pada *Euclidean space*. *Euclidean space* digunakan untuk menganalisis hubungan antara sudut terhadap jarak. *Euclidean Distance* dapat diterapkan dengan menggunakan rumus persamaan (2) [22]:

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - Y_i)^2} \quad (2)$$

dengan:

- x_i : data uji
- Y_i : data nilai

Setelah dilakukan pendeteksian anomali, langkah selanjutnya yang diambil adalah melakukan penanganan data anomali tersebut. Penanganan yang dilakukan pada penelitian ini adalah dengan menghapus data-data anomali yang sudah terdeteksi sebelumnya. Sehingga, setelah penanganan tersebut

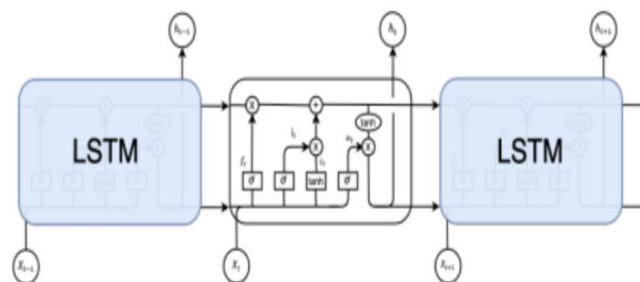
kualitas dataset menjadi lebih baik serta dapat meningkatkan akurasi dan kinerja model dalam melakukan proses data.

Dalam penelitian ini, pemilihan parameter epsilon (ϵ) dan MinPts dilakukan dengan pendekatan *elbow method* untuk menentukan nilai optimal. Metode ini diterapkan dengan menghitung jarak kelima tetangga terdekat dari setiap titik dalam dataset. Selanjutnya, nilai jarak ke-5 terbesar dalam persentil 90% diambil sebagai acuan untuk menentukan nilai ϵ . Pendekatan ini memastikan bahwa nilai ϵ yang dipilih tidak terlalu kecil sehingga membentuk terlalu banyak *cluster* kecil, maupun terlalu besar sehingga menggabungkan *cluster* yang seharusnya terpisah. Dengan cara ini, model DBSCAN dapat lebih akurat dalam mengidentifikasi struktur kepadatan data serta mendeteksi *outlier* secara lebih efektif.

Setelah menentukan nilai ϵ , parameter MinPts juga perlu ditentukan agar dapat mengelompokkan data dengan baik. Dalam penelitian ini, nilai MinPts ditentukan berdasarkan aturan umum, yaitu minimal sebesar dua kali jumlah dimensi data. Namun, karena dataset ini memiliki dimensi yang lebih kompleks, pendekatan berbasis *elbow method* juga diterapkan untuk menyesuaikan nilai MinPts agar lebih optimal. Dengan kombinasi perhitungan ini, DBSCAN dapat mengidentifikasi *cluster* secara lebih adaptif terhadap distribusi kepadatan data, memungkinkan deteksi anomali yang lebih presisi dalam analisis spasial dataset yang digunakan.

2.4. LSTM (Long-Short Term Memory)

Pemodelan atau modelling merupakan langkah yang dapat dilakukan setelah data melalui tahap *preprocessing data*. Prediksi data ril time series merupakan keunggulan dari LSTM [6]. Berbeda dengan algoritma *machine learning* yang lainnya, *Long-Short Term Memory* (LSTM) adalah bagian dari *Recurrent Neural Network* (RNN) [33]. LSTM muncul karena berfungsi untuk mengatasi masalah hilangnya gradien yang menghilang saat menghadapi proses *vanish* dan *exploding* gradien yang terjadi pada RNN [27]. LSTM merupakan algoritma yang diketahui mampu membangun model prediksi dengan baik [34]. LSTM memiliki 3 lapisan yaitu, lapisan input, lapisan tersembunyi, serta lapisan *output* seperti yang dapat dilihat pada Gambar 3.



Gambar 3. LSTM Architecture [35]

Di dalam arsitektur LSTM terdapat yang namanya gerbang atau *gate*. *Gate* tersebut terdapat beberapa jenis, yaitu *forget gate*, *input gate*, *cell state*, *output gate*. *Forget gate* berfungsi sebagai penyaring informasi manakah yang perlu dihapus atau tidak pada *cell state* atau memori jangka panjang. Nilai-nilai yang ada pada *cell state* serta inputan yang memperoleh *forget vector* serta *hidden state* yang diterima akan dialihkan. Untuk *input gate* terdapat 2 fungsi, yaitu dapat melakukan pembaruan informasi yang ada pada *cell state* yang akan diatur oleh *layer sigmoid*, serta membangun kandidat vektor baru untuk meninjau informasi baru yang nantinya akan ditambahkan dengan menggunakan fungsi aktivasi pada *layer tanh*. Nantinya vektor akan digunakan untuk pembaruan kondisi informasi dari *cell state*. *Output gate* memperoleh *hidden state* baru dengan bekerja pada dua *layer*, kemudian akan dilanjutkan ke *timestep* berikutnya. Menentukan informasi yang akan muncul menurut *input* yang ada di waktu ini dengan menggunakan aktivasi sigmoid merupakan pekerjaan yang dilakukan di *layer* pertama.

Sedangkan *layer* yang kedua bertugas dalam penempatan nilai yang terdapat di dalam *cell state* menggunakan aktivasi *tanh* [35]. LSTM dapat dilakukan perhitungan dengan menggunakan rumus-rumus:

$$f_{gt} = \sigma(w_f[o_{t-1}, i_t]) + bs_f \quad (3)$$

$$i_{gt} = \sigma(w_f[o_{t-1}, i_t]) + bs_{igt} \quad (4)$$

$$\hat{C}_{gt} = \tanh(w_f[o_{t-1}, i_t]) + bs_c \quad (5)$$

$$O_{gt} = \sigma(w_f[o_{t-1}, i_t]) + bs_{ogt} \quad (6)$$

$$O_t = O_{gt} \times \tanh(C_t) \quad (7)$$

yang dimana persamaan (3) merupakan persamaan yang diterapkan pada *forget gate*, persamaan (4) dan persamaan (5) diterapkan pada *input gate*, persamaan (6) dan persamaan (7) diterapkan pada bagian *output gate* [28].

Dalam penelitian ini, arsitektur LSTM yang digunakan terdiri dari sejumlah konfigurasi tertentu untuk memastikan performa optimal dalam proses prediksi data time-series. Model LSTM yang dibangun menggunakan 50 neuron pada setiap lapisan LSTM untuk menangkap pola dalam data secara lebih efektif. Aktivasi yang digunakan adalah Tanh pada lapisan LSTM, sementara lapisan dense menggunakan aktivasi linear, karena model ini berfokus pada regresi. Untuk menghindari *overfitting*, diterapkan *dropout* sebesar 20% setelah setiap lapisan LSTM, yang berfungsi untuk mengurangi ketergantungan model terhadap data latih. Proses pelatihan dilakukan dengan 100 *epoch* dan *batch size* sebesar 32, yang memungkinkan model belajar secara bertahap dengan pembaruan bobot yang stabil. Sebagai metode optimasi, digunakan Adam optimizer, yang dikenal efektif dalam pemrosesan data time-series karena kemampuannya dalam menyesuaikan learning rate selama proses pelatihan. Selain itu, diterapkan *Early Stopping* dengan pemantauan terhadap nilai loss, di mana pelatihan akan dihentikan secara otomatis jika dalam 10 *epoch* berturut-turut tidak terjadi peningkatan performa, sehingga dapat menghindari *overfitting* dan meningkatkan efisiensi pelatihan model.

2.5. Analisis Hasil

Pada tahap akhir ini, evaluasi akan dilakukan untuk mengukur tingkat kesalahan (*error*) dari model yang digunakan. Evaluasi ini bertujuan untuk menentukan apakah performa model sudah optimal atau memerlukan perbaikan lebih lanjut. Metode evaluasi yang dipilih adalah *Mean Squared Error* (MSE). MSE dipilih karena kemampuannya untuk memberikan gambaran yang jelas mengenai seberapa jauh prediksi model dari nilai aktual. MSE menghitung rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual, yang membuat metrik ini sangat sensitif terhadap kesalahan besar. Hal ini menjadikan MSE sangat berguna dalam konteks prediksi harga saham, di mana perbedaan kecil antara nilai prediksi dan nilai aktual dapat memiliki dampak yang signifikan terhadap keputusan investasi. Dibandingkan dengan metrik lain seperti *Mean Absolute Error* (MAE) atau *Root Mean Squared Error* (RMSE), MSE lebih memberikan penalti yang lebih besar terhadap kesalahan besar, yang memfokuskan perhatian pada kesalahan yang sangat besar yang perlu diminimalkan dalam model prediksi harga saham. MSE merupakan metrik yang sering digunakan untuk mengukur tingkat kesalahan rata-rata kuadrat antara data ril dan data prediksi. Metrik ini memberikan penilaian terhadap akurasi model dengan memberikan penalti yang lebih besar untuk kesalahan yang lebih besar. Persamaan MSE dapat dituliskan (8) [36]:

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (8)$$

yang dimana:

- x_i : nilai data aktual
- \hat{x}_i : nilai data prediksi yang dihasilkan model

Interpretasi dari nilai MSE yang dihasilkan dapat memberikan informasi yang berguna mengenai performa model. Nilai MSE yang rendah menunjukkan bahwa prediksi harga saham model mendekati nilai aktual dengan kesalahan yang relatif kecil, yang berarti model memiliki kemampuan prediksi yang baik. Sebaliknya, nilai MSE yang tinggi menunjukkan adanya kesalahan prediksi yang signifikan, yang mungkin disebabkan oleh pemilihan parameter yang kurang tepat, model yang kurang kompleks, atau data yang tidak diproses dengan optimal. Jika hasil analisis menunjukkan bahwa nilai MSE masih cukup besar, maka langkah evaluasi lebih lanjut akan dieksekusi. Hal tersebut bertujuan untuk mengidentifikasi faktor-faktor yang menyebabkan performa model kurang memadai, seperti pemilihan parameter model, *preprocessing data*, atau kompleksitas model. Selanjutnya, perbaikan akan diimplementasikan untuk meningkatkan kualitas prediksi.

3. HASIL DAN PEMBAHASAN

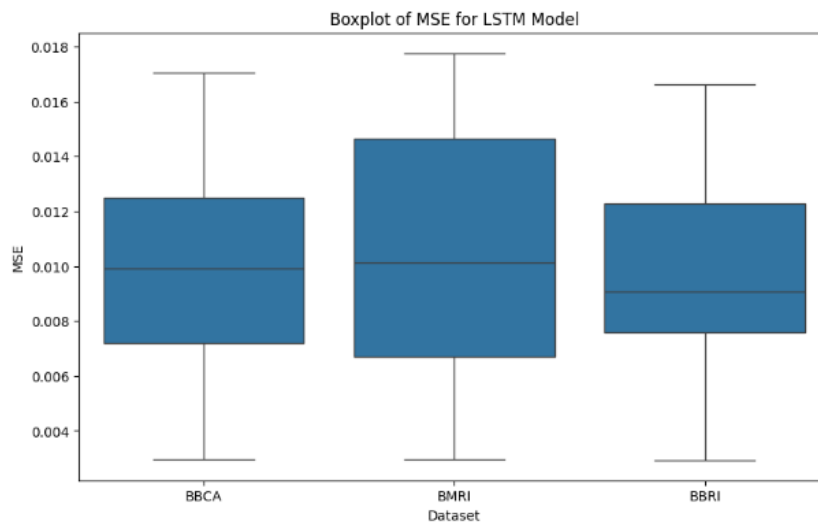
Pendekatan deteksi anomali dengan menggunakan DBSCAN pada model LSTM memungkinkan identifikasi pola anomali dalam data historis secara efektif. Dalam pendekatan ini, DBSCAN digunakan untuk mendeteksi *outlier* yang mungkin mengganggu pola data, sementara LSTM memodelkan urutan data untuk menghasilkan prediksi yang lebih akurat. Integrasi kedua metode ini meningkatkan ketepatan dalam mengidentifikasi perubahan mendadak atau fluktuasi ekstrem yang dapat mempengaruhi performa model. Selain itu, pendekatan ini memberikan solusi yang lebih efisien, baik dari segi pengolahan data maupun akurasi prediksi, untuk menangani data dengan pola tak terduga.

3.1. Skenario Deteksi Anomali

Tabel 4. Comparison of MSE With and Without DBSCAN

K-Fold	MSE					
	DBSCAN			Without DBSCAN		
	BBCA	BMRI	BBRI	BBCA	BMRI	BBRI
1	0.012	0.015	0.012	0.003	0.003	0.004
2	0.003	0.003	0.003	0.006	0.003	0.009
3	0.007	0.007	0.008	0.001	0.002	0.001
4	0.010	0.010	0.009	0.001	0.001	0.001
5	0.017	0.018	0.017	0.000	0.000	0.000

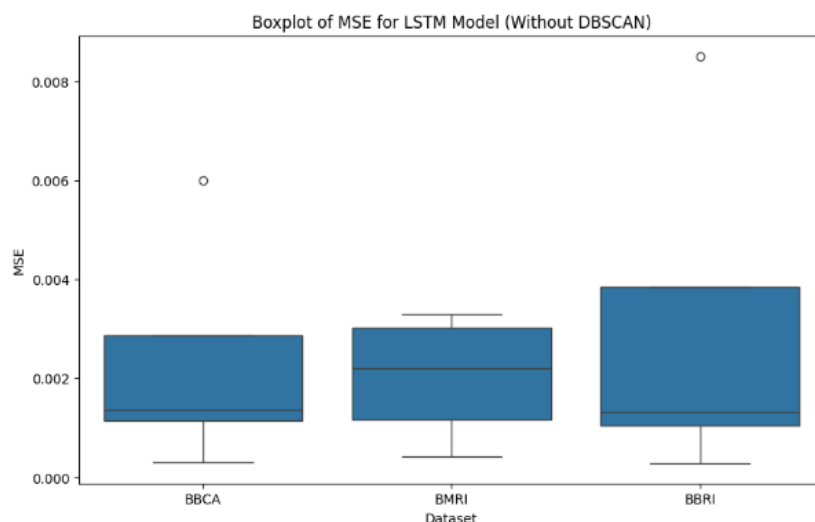
Pada proses ini, deteksi *outlier* dimulai dengan penentuan parameter utama DBSCAN, yaitu epsilon (jarak maksimum antar data yang dianggap sebagai tetangga) dan minPts (jumlah minimum data yang membentuk sebuah kluster). Berdasarkan hasil uji coba dan *tuning parameter*, kami menemukan nilai *epsilon* dan minPts yang optimal untuk mengidentifikasi titik-titik yang tidak sesuai dengan pola umum data. Titik-titik data yang tidak masuk dalam kluster apa pun dikelompokkan sebagai *noise* atau *outlier*. Selanjutnya, data yang telah ditandai sebagai *outlier* dikeluarkan dari dataset sebelum melanjutkan proses analisis lebih lanjut. Setelah proses ini selesai, dataset bersih yang telah melalui tahap deteksi *outlier* ini kemudian digunakan untuk analisis lebih lanjut, termasuk proses prediksi dan evaluasi model. Hasil MSE dengan menggunakan DBSCAN juga dapat dilihat pada Gambar 4.



Gambar 4. MSE Graphic for LSTM with DBSCAN

Dari ketiga saham tersebut terlihat bahwa BBRI memiliki nilai MSE lebih rendah dan lebih konsisten dengan median MSE di angka 0.009, sebaran data juga relatif sempit. Diikuti dengan BBCA dengan sebaran data yang simetris yang memiliki median MSE sebesar 0.010. Sedangkan untuk BMRI terlihat memiliki nilai MSE yang lebih besar diantara ketiganya dan sebaran data yang lebih lebar dengan median MSE sekitar 0.011 - 0.012. Hal tersebut menunjukkan bahwa model LSTM melakukan prediksi pada saham BBRI dengan baik, diikuti oleh BBCA, dan BMRI yang masih kurang akurat.

Sebagai perbandingan, pada penelitian ini juga dilakukan analisis menggunakan dataset yang sama namun tanpa melalui proses deteksi *outlier*. Data ini diolah langsung tanpa tahapan DBSCAN, sehingga semua titik data asli tetap dimasukkan dalam analisis. Langkah ini memungkinkan perbandingan yang komprehensif antara hasil analisis dengan dan tanpa deteksi *outlier*, memberikan wawasan tambahan tentang dampak *outlier* terhadap performa model. Hasil MSE tanpa menggunakan DBSCAN juga dapat dilihat pada Gambar 5.



Gambar 5. MSE Graphic for LSTM without DBSCAN

Berbeda dengan hasil MSE dengan menggunakan DBSCAN, pada hasil ini terlihat bahwa BMRI menunjukkan hasil yang lebih konsisten dibandingkan dengan kedua saham lainnya dengan median MSE sedikit diatas 0.002. Sementara itu, BBCA memiliki nilai median MSE dibawah BMRI yaitu di bawah 0.002 namun terdapat *outlier* di sekitar 0.006 yang menunjukkan prediksi yang dilakukan masih

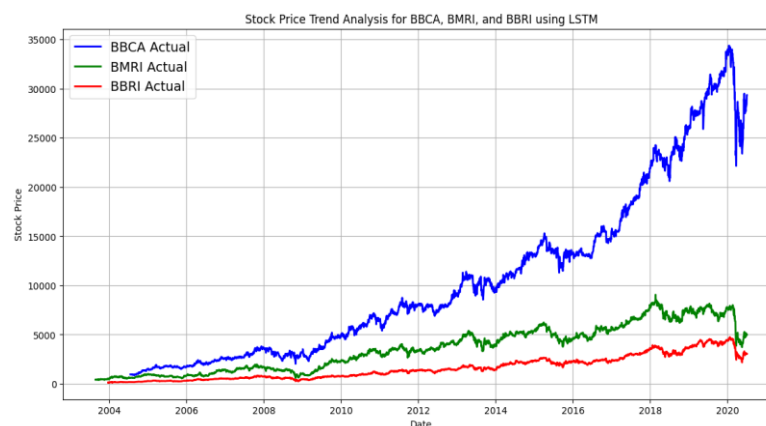
kurang akurat di sebagian titik. Di sisi lain, BBRI memiliki nilai median MSE yang hampir sama dengan BBKA yaitu dibawah 0.002 namun terdapat titik *outlier* di atas 0.008.

Hasil analisis yang tertera pada Tabel 4 menunjukkan adanya perbedaan performa antara model yang menggunakan data dengan deteksi *outlier* melalui DBSCAN dan model tanpa deteksi *outlier*. Model yang dilatih dengan data yang telah dibersihkan menggunakan DBSCAN menunjukkan performa yang lebih baik, dengan nilai *error* yang lebih rendah dan konsistensi prediksi yang lebih tinggi dibandingkan model yang dilatih dengan dataset asli tanpa deteksi *outlier*. Temuan ini menegaskan pentingnya deteksi dan penghapusan *outlier* dalam meningkatkan kualitas prediksi dan stabilitas model.

Uji statistik juga dilakukan pada penelitian ini dengan menggunakan Uji T. Nilai uji-t yang diperoleh adalah -4.345 dengan nilai p-value sebesar 0.001. Nilai p-value memiliki nilai yang lebih kecil dari tingkat signifikansi yang umum digunakan yaitu 0.05. Dari hasil tersebut dapat diketahui bahwa DBSCAN yang diterapkan pada model LSTM dapat meningkatkan kinerja model secara signifikan. Hal ini memiliki implikasi penting untuk investor atau analis pasar saham, bagi investor atau analis pasar saham, penggunaan DBSCAN tetap dapat memberikan manfaat dalam mengurangi risiko akibat data anomali dan menghasilkan prediksi yang lebih akurat dan menjadi lebih stabil dalam pengambilan keputusan investasi.

3.2. Skenario Tren dengan LSTM

Pada skenario tahunan dari harga saham ketiga dataset yang ada yaitu BBKA, BMRI, BBRI dari tahun 2003 hingga 2020 menunjukkan bahwa secara umum ketiganya mengalami kenaikan harga saham selama kurun waktu tersebut walaupun terjadi fluktuasi.



Gambar 6. Trend Graph with LSTM

Pada saham BBRI di tahun 2004-2018 terlihat secara umum tren mengalami kenaikan yang stabil yang menunjukkan bahwa periode tersebut mengalami pertumbuhan yang baik. Pada periode 2018-2020 saham mengalami fluktuasi yang cukup besar, terutama di tahun 2019 yang menunjukkan bahwa terjadinya ketidakpastian pasar ataupun faktor eksternal lainnya yang mempengaruhi saham, dan penurunan yang tajam terjadi di tahun 2020 yang dimana hal ini bisa saja disebabkan oleh pandemi COVID-19. Walaupun sempat terjadi fluktuasi dan penurunan, periode setelah tahun 2020 terlihat terdapat pergerakan naik yang menunjukkan bahwa saham BBRI mulai mengalami adaptasi dan pemulihan. Hasil dari tren tahunan pada saham BBRI dapat dilihat pada Gambar 6.

Untuk saham BMRI setelah dilakukan analisis dari data historis didapatkan pada periode 2004-2018 terjadi peningkatan yang terlihat signifikan, yang dimana hal ini bisa saja ditandai dengan oleh peningkatan daya beli, perkembangan sektor, dan stabilitas ekonomi domestik yang mendorong pertumbuhan harga saham BMRI. Setelah terjadinya kenaikan, saham BMRI sempat mengalami fluktuasi dan volatilitas di periode 2018 hingga 2020. Setelah terjadi puncak kenaikan tertinggi terlihat

penurunan yang lumayan drastis juga. Selain fluktuasi dan volatilitas, harga saham juga sempat mengalami penurunan yang drastis di tahun 2020. Hal ini dikarenakan terjadinya pandemi COVID-19. Pandemi ini berdampak besar pada saham BMRI, sejalan dengan tekanan terhadap stabilitas ekonomi dunia. Pasca 2020 saham BMRI sudah mulai pulih. Hal ini ditunjukkan oleh kenaikan harga saham BMRI yang terjadi pasca 2020. Hasil dari tren tahunan pada saham BMRI dapat dilihat pada Gambar 6.

Sama seperti kedua saham lainnya, saham BBCA juga mengalami kenaikan serta penurunan di beberapa periode. Pertumbuhan jangka panjang terjadi secara signifikan dan konsisten di periode 2004-2018. Fluktuasi juga terjadi di saham BBCA yaitu pada periode 2018-2019 yang terlihat bahwa harga saham BBCA mencapai pada puncaknya yang bisa saja mengindikasikan tingginya permintaan saham BBCA dan diikuti oleh fluktuasi harga yang lebih intens. Penurunan tajam yang disebabkan oleh Pandemi COVID-19 juga terjadi pada saham BBCA. Setelah mengalami penurunan, pasca 2020 saham BBCA mulai menunjukkan pemulihan yang ditandai dengan kenaikan saham. Hasil dari tren tahunan pada saham BBCA dapat dilihat pada Gambar 6.

Dari hasil analisa trend tahunan harga saham BBRI, BMRI, dan BBCA dengan penggunaan LSTM menunjukkan bahwa pada ketiga saham tersebut sama sama terjadi kenaikan yang signifikan terutama periode 2004-2018. Ketiganya juga sama sama terjadi fluktuasi serta penurunan yang drastis di tahun 2020 yang disebabkan oleh pandemi COVID-19. Pasca 2020 ketiga saham mengalami pemulihan yang ditandai dengan kenaikan harga saham. Hasil prediksi tersebut dapat menjadi pengetahuan baru untuk para investor sebelum mengambil keputusan agar tidak mengalami kerugian dan tetap mendapatkan keuntungan

4. HASIL DISKUSI

Berbagai penelitian sebelumnya telah menerapkan sejumlah metode deteksi anomali, seperti *Local Outlier Factor* (LOF), DBSCAN, dan *Isolation Forest*. Salah satu studi yang menggunakan DBSCAN melaporkan bahwa sekitar 5% data teridentifikasi sebagai *outlier* [22]. Namun, penelitian tersebut juga menyoroti keterbatasan metode ini dalam menangkap informasi implisit yang memerlukan interpretasi lebih lanjut dari ahli domain. Dalam konteks penelitian ini, penggunaan DBSCAN bertujuan untuk penghapusan *outlier* sehingga dapat meningkatkan kualitas data sebelum dilakukan prediksi harga saham.

Di sisi lain, dalam analisis prediksi harga saham, model *Long Short-Term Memory* (LSTM) telah terbukti efektif dalam menangani data yang bersifat non-linear serta mengenali pola tren jangka panjang. Beberapa penelitian menunjukkan bahwa LSTM memiliki kinerja *superior* dibandingkan metode lainnya dalam menganalisis data saham multi-periode. Sebagai ilustrasi, studi yang menerapkan LSTM untuk memprediksi harga saham BBRI mencatat nilai RMSE sebesar 227,47 [24]. Sementara itu, penelitian lain yang meneliti dampak pandemi Covid-19 terhadap prediksi saham menggunakan *deep learning* melaporkan hasil RMSE yang bervariasi, yakni BBCA (1120,65), BBRI (156,29), ASII (134,55), dan TLKM (71,65) [25].

Evaluasi yang dilakukan dalam penelitian ini mengungkapkan bahwa kombinasi model LSTM dengan DBSCAN menghasilkan nilai *Mean Squared Error* (MSE) yang lebih rendah dibandingkan skenario tanpa penghapusan *outlier*. Pada saham BBCA, MSE terendah tercatat 0.003 pada fold ke-2 dengan DBSCAN, sedangkan tanpa DBSCAN mencapai 0.006. Pada saham BMRI, MSE terbaik sebesar 0.003 tercapai pada fold ke-4 dengan DBSCAN, sedangkan tanpa DBSCAN pada fold ke-5 mencapai 0.000. Untuk saham BBRI, MSE terbaik 0.003 tercatat pada fold ke-2 dengan DBSCAN, dan fold ke-5 tanpa DBSCAN sebesar 0.000. Dengan demikian, LSTM dengan DBSCAN pada fold ke-2 memberikan prediksi yang lebih akurat.

Tabel 5. Comparison Results of Previous Research with This Research

Aspect	Previous Research		This Research
	A. S. B. Karno. [2020]	W. Hastomo, et al. [2021]	
Anomaly Detection Datasets	There isn't any Bank BRI	There isn't any BBCA, BBRI, ASII, TLKM	DBSCAN BBCA, BMRI, BBRI
Model Evaluation Accuracy	RMSE 227,47033324 4533	RMSE BBCA (1120.65), BBRI (156.29), ASII (134.55), dan TLKM (71.65)	MSE BBCA (0.003), BMRI (0.003), BBRI (0.003)

Berdasarkan hasil perbandingan dengan penelitian sebelumnya Tabel 5 penghapusan *outlier* dengan DBSCAN berdampak signifikan pada saham dengan volatilitas tinggi seperti BBCA, tetapi tidak memberikan perubahan besar pada saham yang lebih stabil seperti BMRI dan BBRI. Hal ini menunjukkan bahwa pengaruh *outlier* bergantung pada karakteristik volatilitas saham. Penelitian ini berbeda dari penelitian sebelumnya yang lebih berfokus pada identifikasi *outlier* tanpa mengintegrasikan langkah penghapusan sebelum prediksi. Sementara penelitian terdahulu melaporkan sekitar 5% data sebagai *outlier*, penelitian ini menyoroti bahwa efektivitas penghapusan *outlier* tergantung pada parameter DBSCAN, karakteristik dataset, serta model prediksi yang digunakan.

Namun, penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Salah satu keterbatasan utama adalah asumsi bahwa semua *outlier* yang terdeteksi oleh DBSCAN harus dihapus. Namun dalam beberapa kasus, *outlier* dapat mengandung informasi penting mengenai pergerakan harga saham yang tidak dapat diabaikan begitu saja. Selain itu, penelitian ini belum mengeksplorasi lebih lanjut pemilihan parameter optimal untuk DBSCAN guna memastikan bahwa proses deteksi anomali dilakukan secara efektif untuk setiap jenis saham yang dianalisis. Keterbatasan lainnya adalah rentang waktu dan cakupan dataset yang digunakan, sehingga hasil penelitian ini mungkin tidak dapat digeneralisasi untuk semua kondisi pasar saham. Oleh karena itu, penelitian lanjutan dapat mempertimbangkan eksperimen dengan parameter DBSCAN yang lebih luas atau bahkan membandingkannya dengan metode deteksi anomali lainnya, seperti *Isolation Forest* atau metode lainnya, guna mengevaluasi dampak dari berbagai pendekatan dalam meningkatkan akurasi prediksi harga saham.

Implikasi dari penelitian ini terhadap praktik prediksi harga saham adalah bahwa penggunaan metode deteksi anomali sebagai langkah *preprocessing* dapat meningkatkan akurasi prediksi, tetapi efektivitasnya sangat bergantung pada karakteristik saham yang dianalisis. Untuk saham dengan volatilitas tinggi, penghapusan *outlier* terbukti membantu meningkatkan akurasi model, tetapi untuk saham yang lebih stabil, pendekatan ini mungkin tidak memberikan manfaat yang signifikan. Oleh karena itu, dalam implementasi nyata, investor dan analis pasar saham perlu mempertimbangkan sifat pergerakan harga saham sebelum menerapkan teknik deteksi dan penghapusan *outlier*. Di masa mendatang penelitian ini dapat dikembangkan dengan menguji pendekatan hybrid yang menggabungkan metode deteksi anomali dengan teknik pembobotan atau seleksi fitur, sehingga *outlier* yang masih memiliki informasi berharga tidak serta-merta dihapus, tetapi dianalisis lebih lanjut untuk meningkatkan akurasi model prediksi harga saham.

5. KESIMPULAN

Pada saham BBCA, model LSTM dengan penghapusan *outlier* menggunakan DBSCAN mencatat MSE terendah sebesar 0.003 pada fold ke-2, dibandingkan dengan tanpa DBSCAN yang mencapai nilai 0.006. Untuk saham BMRI, hasil terbaik diperoleh pada fold ke-2 dengan MSE sebesar 0.003 dengan menggunakan DBSCAN dan fold ke-5 tanpa DBSCAN sebesar 0.000. Sementara itu, pada saham BBRI, model LSTM juga menunjukkan MSE terbaik sebesar 0.003 pada fold ke 2 dengan penggunaan DBSCAN dan fold ke-5 tanpa menggunakan DBSCAN sebesar 0.000. Berdasarkan temuan ini, dapat disimpulkan bahwa penggunaan DBSCAN dalam proses penghapusan outlier dapat menghasilkan peningkatan akurasi prediksi dalam beberapa kasus, tetapi tidak selalu memberikan hasil yang lebih baik dibandingkan dengan model tanpa DBSCAN. Pada fold ke-5, misalnya, semua saham memiliki nilai MSE 0.000 tanpa DBSCAN, menunjukkan bahwa dalam beberapa kondisi, model tanpa penghapusan outlier dapat memberikan prediksi yang lebih akurat. Oleh karena itu, metode penanganan outlier selain DBSCAN, seperti Local Outlier Factor (LOF) atau Isolation Forest, dapat dieksplorasi dalam penelitian selanjutnya untuk membandingkan efektivitasnya dalam meningkatkan performa model LSTM. Selain itu, penelitian dapat diperluas juga dengan penambahan dataset terbaru.

DAFTAR PUSTAKA

- [1] D. Tambunan, "Investasi Saham di Masa Pandemi COVID-19," *Widya Cipta J. Sekr. dan Manaj.*, vol. 4, no. 2, pp. 117–123, 2020.
- [2] T. Febiyola, R. S. Utari, B. T. Panggabean, and R. Agustina, "Analisis Surat Berharga Sebagai Alat Investasi," *Publ. Ilmu Huk.*, vol. 2, no. 3, pp. 75–86, 2024.
- [3] M. Azizah, M. I. Irawan, and E. R. M. Putri, "Comparison of stock price prediction using geometric Brownian motion and multilayer perceptron," *AIP Conf. Proc.*, vol. 2242, no. April, 2020.
- [4] W. C. Utomo, "Prediksi Pergerakan Saham BBRI ditengah Issue Ancaman Resesi 2023 dengan Pendekatan Machine Learning," *J. Teknol. dan Manaj. Inform.*, vol. 9, no. 1, pp. 20–27, 2023.
- [5] S. Dwi Ceysa, J. Demar Putri, D. A. Putri, and F. Siswajanthy, "Peranan Perbankan dalam Perekonomian Indonesia," vol. 8, pp. 25959–25964, 2024.
- [6] W. Budiharto, "Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM)," *J. Big Data*, vol. 8, no. 1, 2021.
- [7] R. S. Lorenza, R. Octavia, S. Shafitranata, and A. W. Madyoningrum, "Dampak Covid 19 Terhadap Perubahan Harga Saham Perbankan di Indonesia Sebelum dan Saat Pandemi," *J. Akuntansi, Keuangan, dan Manaj.*, vol. 4, no. 1, pp. 43–56, 2022.
- [8] A. M. Priyatno, L. S. Tanjung, W. F. Ramadhan, P. Cholidhazia, P. Z. Jati, and F. I. Firmananda, "Comparison Random Forest Regression and Linear Regression For Forecasting BBCA Stock Price," *J. Tek. Ind. Terintegrasi*, vol. 6, no. 3, pp. 718–732, 2023.
- [9] Bursa Efek Indonesia, "50 Biggest Market Capitalization - Juli 2024," 2024. [Online]. Available: <https://idx.co.id/id/data-pasar/laporan-statistik/digital-statistic/monthly/biggest-market-capitalization-most-active-stocks/biggest-market-capitalization?filter=eyJ5ZWFyIjoiMjAyMyIsImIvbnRlIjoiMyIsInF1YXJ0ZXIiOiJAsInR5cGUiOiJtb250aGx5In0%3D>.
- [10] P. R. Sihombing, S. Suryadiningrat, D. A. Sunarjo, and Y. P. A. C. Yuda, "Identifikasi Data *Outlier* (Pencilan) dan Kenormalan Data Pada Data Univariat serta Alternatif Penyelesaiannya," *J. Ekon. Dan Stat. Indones.*, vol. 2, no. 3, pp. 307–316, 2023.
- [11] K. Choi, J. Yi, C. Park, and S. Yoon, "Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines," *IEEE Access*, vol. 9, pp. 120043–120065, 2021.
- [12] T. Bank, "Jurnal Teknologi Terpadu DIFFERENTIAL EVOLUTION UNTUK DETEKSI ANOMALI PADA DATA," vol. 10, no. 1, pp. 22–31, 2024.
- [13] N. Nasution and F. Rakhmawati, "Segmentasi Pengguna E-Wallet Dengan Menggunakan Metode Dbscan (Density Based Spatial Clustering Application With Noise) Di Kota Medan," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 4, no. 2, pp. 1386–1392, 2023.

-
- [14] P. O. C. Sari and S. Suharjito, "Outlier Detection in Inpatient Claims Using DBSCAN and K-Means," *J. Tek. Inform.*, vol. 15, no. 1, pp. 1–10, 2022.
- [15] Y. Syawali, M. Haikal, H. Rangkuti, and K. A. Mayadi, "PENERAPAN ALGORITMA DBSCAN UNTUK ANALISIS DEMOGRAFIS dan PENGELUARAN PELANGGAN MALL."
- [16] D. Armiady, "Analisis Metode DBSCAN (Density-Based Spatial Clustering of Application with Noise) dalam Mendeteksi Data Outlier," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 6, p. 2158, 2022.
- [17] R. Saputra *et al.*, "ANALISIS PREDIKSI SAHAM TESLA MENGGUNAKAN ALGORITMA LONG SHORT TERM MEMORY (LSTM)," vol. 2, no. 1, pp. 81–90, 2024.
- [18] V. Arinal and M. Puspita, "Peningkatan Akurasi Nilai Harga Saham Menggunakan Metode Long Short-Term Memory (LSTM) pada PT Unilever Abstrak," vol. 6, no. 1, pp. 252–260, 2025.
- [19] R. F. Inaku and J. C. Chandra, "Implementasi Data Mining Dalam Prediksi Harga Saham Menggunakan Metode Long Short Term Memory (Lstm)," *J. Ticom Technol. Inf. Commun.*, vol. 12, no. 1, pp. 1–7, 2023.
- [20] R. Luthfiansyah and B. Wasito, "Penerapan Teknik Deep Learning (Long Short Term Memory) dan Pendekatan Klasik (Regresi Linier) dalam Prediksi Pergerakan Saham BRI," *J. Inform. dan Bisnis*, vol. 12, no. 2, pp. 42–54, 2023.
- [21] R. Julian and M. R. Pribadi, "Peramalan Harga Saham Pertambangan Pada Bursa Efek Indonesia (BEI) Menggunakan Long Short Term Memory (LSTM)," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 3, pp. 1570–1580, 2021.
- [22] E. A. Fadlilah, "Identifikasi Anomali Data Akademik Menggunakan Dbscan Outlier Detection," *Pros. Sains Nas. dan Teknol.*, vol. 12, no. 1, p. 336, 2022.
- [23] A. Primawati, I. S. Sitanggang, A. Annisa, and D. A. Astuti, "Perbandingan Kinerja LSTM dan Prophet untuk Prediksi Deret Waktu (Studi Kasus Produksi Susu Sapi Harian)," *J. Edukasi dan Penelit. Inform.*, vol. 9, no. 3, p. 428, 2023.
- [24] A. S. B. Karno, "Prediksi Data Time Series Saham Bank BRI Dengan Mesin Belajar LSTM (Long ShortTerm Memory)," *J. Inform. Inf. Secur.*, vol. 1, no. 1, pp. 1–8, 2020.
- [25] W. Hastomo, A. S. B. Karno, N. Kalbuana, E. Nisfiani, and L. ETP, "Optimasi Deep Learning untuk Prediksi Saham di Masa Pandemi Covid-19," ... (*Jurnal Edukasi dan ...*, vol. 7, no. 2, pp. 133–140, 2021).
- [26] Bursa Efek Indonesia, "Indeks Saham." [Online]. Available: <https://www.idx.co.id/id/data-pasar/data-saham/indeks-saham>.
- [27] N. P. A. Widiari, I. M. A. D. Suarjaya, and D. P. Githa, "Teknik Data Cleaning Menggunakan Snowflake untuk Studi Kasus Objek Pariwisata di Bali," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 8, no. 2, p. 137, 2020.
- [28] M. R. A. Prasetya, A. M. Priyatno, and Nurhaeni, "Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining," *J. Inf. dan Teknol.*, vol. 5, no. 2, pp. 52–62, 2023.
- [29] Z. Efendy, "Normalisasi dalam Desain Database," *J. CoreIT*, vol. 4, no. 1, pp. 34–43, 2018.
- [30] P. J. Muhammad Ali, "Investigating the Impact of Min-Max Data Normalization on the Regression Performance of K-Nearest Neighbor with Different Similarity Measurements," *Arothe Sci. J. Koya Univ.*, vol. 10, no. 1, pp. 85–91, 2022.
- [31] I. D. Id, A. Astrid, and E. Mahdiyah, "Modifikasi DBSCAN (Density-Based Spatial Clustering With Noise) pada Objek 3 Dimensi," *J. Komput. Terap.*, vol. 3, no. 1, pp. 41–52, 2017.
- [32] F. Izhari, "Analisis Algoritma Dbscan Dalam Menentukan Parameter Epsilon Pada Clustering Data Numerik," *Semin. Nas. Teknol. Komput. Sains*, pp. 156–158, 2020.
- [33] M. A. Maliki, I. Cholissodin, and N. Yudistira, "Prediksi Pergerakan Harga Cryptocurrency Bitcoin terhadap Mata Uang Rupiah menggunakan Algoritme LSTM," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 7, pp. 3259–3268, 2022.
- [34] J. Asbullah and S. Samsudin, "Prediksi Harga Cryptocurrency Binance Berdasarkan Informasi Blockchain dengan Menggunakan Algoritma Random Forest," *J. Media Inform. Budidarma*, vol. 8, no. 1, p. 260, 2024.
- [35] D. R. Chandranegara, R. A. Afif, C. S. K. Aditya, W. Suharso, and H. Wibowo, "Prediksi Harga
-

Saham Jakarta Islamic Index Menggunakan Metode Long Short-Term Memory,” *J. Edukasi dan Penelit. Inform.*, vol. 9, no. 1, p. 129, 2023.

[36] A. Rosenblad, *The Concise Encyclopedia of Statistics*, vol. 38, no. 4. 2011.