

ANALISIS POLA DAN PERGERAKAN HARGA SAHAM NVIDIA DENGAN KLASTERISASI K-MEANS DAN PREDIKSI TIME-SERIES

Hansel Valent, Fhadillah, Rosyid Fauzan, Nazwar Farezi, Fanny Ramadhani

Ilmu Komputer, Universitas Negeri Medan

Jalan Williem Iskandar Pasar V, Medan, Sumatera Utara, Indonesia

hanseldragon557@gmail.com

ABSTRAK

Pasar saham merupakan indikator penting dalam ekonomi, dengan harga saham yang fluktuatif mencerminkan berbagai faktor seperti sentimen pasar dan kondisi perusahaan. NVIDIA Corporation (NVDA) menjadi salah satu perusahaan teknologi yang menarik perhatian investor karena pertumbuhannya yang signifikan dalam industri semikonduktor dan AI. Namun, volatilitas tinggi menjadi tantangan dalam menentukan strategi investasi. Penelitian ini bertujuan untuk menganalisis pola pergerakan harga saham NVIDIA menggunakan klusterisasi K-Means dan memprediksi harga saham dengan model Long Short-Term Memory (LSTM). Permasalahan utama adalah kesulitan model prediksi dalam lonjakan harga ekstrem dan ketergantungan pada data historis. Tujuan penelitian adalah mengidentifikasi pola harga saham melalui klusterisasi dan menguji akurasi prediksi LSTM dengan berbagai skenario data latih. Metode yang digunakan meliputi pengumpulan data historis NVIDIA (2020–2025) dari Yahoo Finance, preprocessing data, klusterisasi K-Means dengan penentuan kluster optimal menggunakan metode Elbow, serta prediksi time-series dengan LSTM. Evaluasi dilakukan menggunakan RMSE dan MAPE. Hasil klusterisasi menunjukkan tiga kluster optimal dengan Silhouette Score 0.579, mengindikasikan pengelompokan yang cukup baik. Prediksi LSTM menghasilkan RMSE dan MAPE terendah pada 60% data latih (70.18 dan 9.61%), namun kesalahan meningkat seiring penambahan data latih akibat ketidakmampuan model menangkap lonjakan harga ekstrem. Disarankan untuk menambahkan fitur seperti volume perdagangan guna meningkatkan akurasi prediksi.

Kata kunci : *Klusterisasi, K-Means, LSTM, Prediksi Saham, NVIDIA.*

1. PENDAHULUAN

Pasar saham merupakan salah satu indikator utama dalam menggambarkan kondisi ekonomi suatu negara maupun kinerja perusahaan secara spesifik. Harga saham yang berfluktuasi setiap hari mencerminkan sentimen pasar, kondisi fundamental perusahaan, serta faktor eksternal seperti kebijakan ekonomi, inflasi, dan peristiwa global. Dalam beberapa tahun terakhir, NVIDIA Corporation (NVDA) telah menjadi salah satu perusahaan teknologi yang menarik perhatian investor, terutama karena perannya dalam industri semikonduktor, kecerdasan buatan (AI), dan komputasi berbasis GPU. Saham NVIDIA mengalami pertumbuhan yang signifikan dalam satu dekade terakhir, dengan kenaikan harga lebih dari 700% sejak tahun 2015. Namun, volatilitas yang tinggi juga menjadi tantangan bagi investor dalam menentukan strategi investasi yang optimal.

Dalam analisis pasar saham, pemanfaatan teknik data mining dan machine learning menjadi semakin penting untuk memahami pola pergerakan harga saham dan membuat prediksi yang lebih akurat. Salah satu pendekatan yang digunakan adalah klusterisasi, di mana data harga saham dikelompokkan berdasarkan pola pergerakan tertentu menggunakan metode seperti K-Means. Metode ini memungkinkan pengelompokan saham berdasarkan kesamaan tren, yang dapat membantu investor dalam mengidentifikasi periode dengan volatilitas tinggi atau rendah. Selain itu, untuk memahami tren harga saham ke depan, metode

prediksi time-series seperti LSTM dapat digunakan untuk menganalisis pola historis dan memproyeksikan pergerakan harga di masa mendatang.

Meskipun berbagai metode telah diterapkan dalam analisis pasar saham, masih terdapat beberapa tantangan yang perlu diselesaikan. Salah satunya adalah isu overfitting dalam model prediksi yang menyebabkan hasil tidak akurat saat diterapkan pada data baru. Selain itu, data saham sering kali dipengaruhi oleh faktor eksternal yang sulit dimodelkan, seperti kebijakan makroekonomi atau sentimen pasar global. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan pendekatan analisis yang lebih efektif dengan menggunakan K-MEANS untuk memahami pola pergerakan harga saham NVIDIA, serta menerapkan model prediksi time-series untuk memperkirakan tren harga di masa depan.

Dalam penelitian ini, analisis dilakukan dengan membandingkan pola pergerakan harga saham NVIDIA selama periode 2020–2025. Data historis yang digunakan berasal dari sumber terpercaya seperti Yahoo Finance, yang mencakup informasi harga pembukaan, penutupan, volume perdagangan, dan volatilitas harian. Dengan mengintegrasikan metode klusterisasi dan prediksi time-series, penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam mengenai pola pergerakan saham NVIDIA serta memberikan informasi yang berguna bagi investor dalam mengambil keputusan investasi. Selain itu, perbedaan utama dalam penelitian ini dibandingkan dengan studi sebelumnya terletak pada

kombinasi metode yang digunakan serta analisis tren saham NVIDIA secara spesifik, yang relevan mengingat perkembangan industri teknologi dan semikonduktor dalam beberapa tahun terakhir.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Dalam penelitian Hartanti [1], metode Elbow digunakan untuk menentukan jumlah cluster yang ideal untuk mengatur kesiapan siswa SMK untuk Ujian Nasional. Studi ini menemukan bahwa tiga cluster adalah jumlah yang ideal setelah membandingkan nilai Sum of Square Error (SSE) dari berbagai jumlah cluster. Hasil pengujian menunjukkan bahwa teknik Elbow dapat mengsegmentasi data dengan lebih baik untuk membedakan tingkat kesiapan siswa berdasarkan nilai akademik mereka. Penelitian ini juga menekankan betapa pentingnya memilih jumlah kelompok yang tepat untuk meningkatkan efisiensi analisis data di bidang pendidikan.

Penelitian yang dilakukan N. H. Harani et al. [2], metode Elbow digunakan untuk menentukan jumlah cluster terbaik dalam segmentasi pelanggan Indihome. Berdasarkan hasil analisis, ditemukan bahwa metode Elbow mampu menghasilkan jumlah cluster optimal yang mendukung peningkatan strategi pemasaran berbasis data. Dengan menggunakan teknik ini, perusahaan dapat mengelompokkan pelanggan berdasarkan pola penggunaan layanan digital mereka, sehingga strategi pemasaran dapat disesuaikan dengan kebutuhan setiap segmen pelanggan.

Penelitian yang dilakukan Gilland Achyar dan Onoy Rohaeni [3] mereka menggabungkan metode K-Means dan General Regression Neural Network (GRNN) untuk memprediksi harga saham pada indeks LQ45. Meskipun tidak secara langsung menggunakan LSTM, pendekatan ini menunjukkan bagaimana klusterisasi dapat meningkatkan akurasi model prediksi harga saham. Dengan mengelompokkan data berdasarkan karakteristik serupa menggunakan K-Means, GRNN dapat mempelajari pola dalam setiap kelompok dengan lebih efektif, sehingga mengurangi kompleksitas jaringan dan komputasi. Hasil penelitian menunjukkan bahwa model hybrid K-Means & GRNN memiliki nilai Mean Absolute Percentage Error (MAPE) sebesar 0,943%, yang menunjukkan akurasi prediksi yang tinggi. Pendekatan ini dapat menjadi inspirasi untuk menggabungkan LSTM dengan teknik klusterisasi seperti K-Means dalam analisis pola dan pergerakan harga saham NVIDIA.

Penelitian yang dibuat Zhang et al. [4] mengusulkan model prediksi harga saham yang menggabungkan Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM). CNN digunakan untuk mengekstraksi fitur spasial dari data harga saham, sementara LSTM digunakan untuk menangkap dependensi temporal. Hasil eksperimen menunjukkan bahwa model gabungan ini memiliki akurasi prediksi yang lebih tinggi dibandingkan

dengan model individu, menunjukkan potensi kombinasi CNN dan LSTM dalam analisis pergerakan harga saham.

2.2. Algoritma K-MEANS

Salah satu metode pembelajaran tanpa pengawasan adalah algoritma K-Means, yang digunakan untuk mengelompokkan data ke dalam berbagai cluster berdasarkan kemiripan atribut. Prosesnya dimulai dengan menentukan jumlah cluster k , kemudian algoritma akan menempatkan centroid awal secara acak dan mengelompokkan data berdasarkan centroid terdekat. Proses ini berulang kali dilakukan hingga konvergensi, ketika tidak ada perubahan signifikan dalam posisi centroid [5].

Salah satu kelebihan utama algoritma K-Means adalah kemampuan untuk menangani dataset dalam jumlah besar dengan komputasi yang relatif cepat. Namun, masalah utama algoritma ini adalah menemukan jumlah k yang optimal, yang biasanya ditentukan menggunakan metode seperti Elbow Method atau Silhouette Score. Selain itu, K-Means seringkali tidak efektif dalam menangani data dengan distribusi bola yang tidak berbentuk atau banyak outlier, sehingga sering dikompromikan [6].

Salah satu penerapan algoritma K-Means yang cukup luas adalah dalam analisis tren dan pola pergerakan harga saham. Dalam konteks ini, K-Means dapat digunakan untuk mengelompokkan pergerakan harga saham berdasarkan karakteristik seperti harga pembukaan (Open), harga penutupan (Close), harga tertinggi (High), harga terendah (Low), serta volume perdagangan. Dengan membagi harga saham ke dalam beberapa kategori, investor dapat memahami pola yang terbentuk dan membuat strategi investasi yang lebih baik. Namun, karena K-Means tidak selalu optimal dalam mendeteksi cluster dengan bentuk yang kompleks, metode ini sering dipadukan dengan algoritma lain seperti DBSCAN untuk meningkatkan akurasi. Dalam penelitian ini, K-Means digunakan untuk mengelompokkan pola pergerakan harga saham NVIDIA, dengan evaluasi hasil clustering dilakukan menggunakan Silhouette Score untuk memastikan kualitas pengelompokan yang dihasilkan [7].

2.3. Metode Elbow

Metode Elbow adalah teknik yang digunakan untuk menentukan jumlah kluster optimal dalam algoritma K-Means dengan menganalisis nilai Sum of Square Error (SSE) pada berbagai jumlah kluster. Titik optimal ditandai oleh "elbow point" pada grafik, yaitu saat penurunan SSE mulai melambat secara signifikan, menunjukkan bahwa penambahan kluster berikutnya tidak memberikan peningkatan berarti terhadap kualitas model. Dalam penelitian oleh Adella Putri Riani, Apriade Voutama, dan Taufik Ridwan terkait pengelompokan hasil belajar peserta didik di PKBM Sanggar Puri, metode ini berhasil mengidentifikasi tiga kluster sebagai jumlah optimal. Penggunaan metode Elbow dalam konteks ini tidak hanya

menghasilkan pengelompokan data yang lebih akurat dan representatif, tetapi juga memperkuat evaluasi performa model melalui pengukuran Davies Bouldin Index. Dengan demikian, pendekatan ini menjadi dasar yang kuat dalam menyederhanakan kompleksitas analisis data sekaligus mengarahkan strategi pembelajaran yang lebih efektif di lingkungan pendidikan [8].

2.4. Long Short-Term Memory

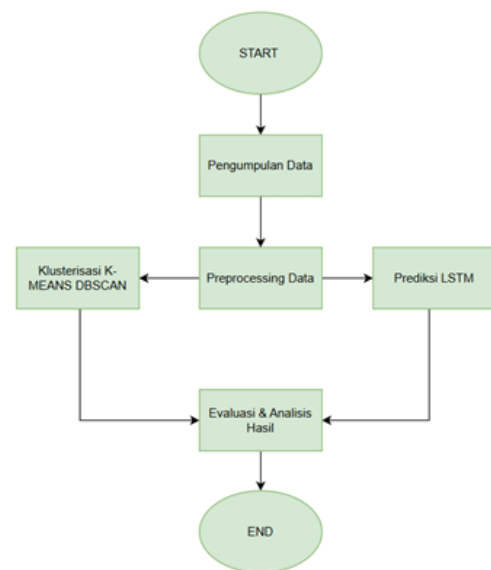
Model Long Short-Term Memory (LSTM) merupakan bentuk lanjutan dari arsitektur Recurrent Neural Network (RNN) yang dirancang untuk mengatasi keterbatasan dalam menangkap dependensi jangka panjang pada data deret waktu. Salah satu keunggulan utama LSTM dibandingkan RNN konvensional adalah kemampuannya mengatasi permasalahan vanishing gradient melalui penerapan mekanisme gate—yakni forget gate, input gate, dan output gate. Ketiga komponen ini bekerja secara sinergis untuk mengatur proses penyimpanan, pembaruan, dan pengeluaran informasi di dalam sel memori, sehingga memungkinkan model untuk mempertahankan informasi relevan dari data historis dalam jangka waktu yang lebih panjang. Keandalan LSTM dalam mengenali pola-pola kompleks menjadikannya populer dalam berbagai aplikasi, seperti prediksi harga saham, analisis iklim, dan peramalan tren ekonomi. Dalam berbagai penelitian, LSTM terbukti memberikan hasil prediksi yang lebih akurat dibandingkan model statistik klasik seperti ARIMA, terutama ketika diterapkan pada dataset yang berukuran besar. Kemampuan LSTM untuk mengelola informasi sekuensial dalam jangka panjang membuatnya sangat sesuai untuk memodelkan data yang bersifat dinamis dan fluktuatif, seperti pasar saham [9].

Dalam konteks prediksi harga saham, LSTM digunakan untuk memperkirakan pergerakan harga di masa mendatang dengan memanfaatkan pola-pola yang terdapat dalam data historis. Tahapan pelatihan model LSTM mencakup proses penting seperti normalisasi data untuk memastikan skala yang seragam, pembagian dataset menjadi data pelatihan (training) dan pengujian (testing) untuk menghindari overfitting, serta penyetelan hyperparameter seperti jumlah epoch, batch size, dan jumlah neuron pada lapisan tersembunyi. Walaupun LSTM memiliki keunggulan dalam menangani karakteristik pasar saham yang volatil dan kompleks, model ini tetap menghadapi tantangan dalam menemukan kombinasi parameter yang optimal guna mencapai hasil prediksi terbaik. Tuning parameter yang kurang tepat dapat berdampak signifikan terhadap performa model. Dalam penelitian ini, model LSTM diterapkan untuk memprediksi harga saham NVIDIA dengan menggunakan data historis dari tahun 2020 hingga 2025. Evaluasi performa dilakukan menggunakan dua metrik utama, yaitu Root Mean Squared Error (RMSE) dan Mean Absolute Percentage Error (MAPE), guna

mengukur tingkat kesalahan dan akurasi dari hasil prediksi yang dihasilkan oleh model. Pendekatan ini diharapkan mampu memberikan gambaran yang lebih akurat mengenai arah pergerakan harga saham di masa depan. [10].

3. METODE PENELITIAN

Dalam penelitian ini, dilakukan analisis data harga saham NVIDIA menggunakan teknik klusterisasi K-Means serta prediksi time-series dengan LSTM. Metode penelitian yang diterapkan mengikuti beberapa tahapan utama, yaitu pengumpulan data, preprocessing, klusterisasi, prediksi, dan evaluasi hasil. Untuk memudahkan pemahaman alur penelitian, keterangan lebih lanjut dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

3.1. Pengumpulan Data

Tahap pertama dalam penelitian ini adalah pengumpulan data historis harga saham NVIDIA yang diperoleh dari Yahoo Finance dengan rentang waktu 2020 hingga 2025 yang di bantu melalui pemograman python. Data yang dikumpulkan mencakup beberapa variabel utama, yaitu tanggal transaksi (Date), harga pembukaan (Open), harga penutupan (Close), harga tertinggi (High), harga terendah (Low) dan volume perdagangan (Volume). Variabel-variabel ini memiliki peran penting dalam analisis pergerakan harga saham dan dapat digunakan untuk mengidentifikasi pola yang terjadi dalam pasar. Data yang telah dikumpulkan kemudian disimpan dalam format CSV (Comma-Separated Values) untuk mempermudah pemrosesan lebih lanjut dalam tahapan analisis berikutnya.

3.2. Preprocessing Data

Pada tahap ini, data yang telah dikumpulkan diperiksa dan diproses agar siap digunakan dalam analisis. Proses preprocessing dilakukan untuk memastikan bahwa data tidak mengandung kesalahan atau nilai yang hilang yang dapat mempengaruhi hasil

analisis. Langkah pertama yang dilakukan adalah pengecekan missing values, di mana jika terdapat data yang hilang, maka dilakukan penghapusan atau pengisian ulang menggunakan metode interpolasi atau imputasi. Selanjutnya, dilakukan identifikasi outlier menggunakan metode visualisasi seperti boxplot untuk mendeteksi adanya anomali dalam pergerakan harga saham. Outlier yang teridentifikasi kemudian dianalisis apakah perlu dihapus atau tetap dipertahankan sebagai bagian dari pola pergerakan saham.

3.3. Klasterisasi dengan K-Means

Tahap selanjutnya adalah klasterisasi menggunakan K-Means untuk memahami pola pergerakan harga saham NVIDIA. K-Means merupakan algoritma clustering berbasis partisi yang mengelompokkan data ke dalam sejumlah k cluster berdasarkan kemiripan pola harga saham, dengan jumlah cluster optimal ditentukan menggunakan Elbow Method dan Silhouette Score.

3.4. Prediksi Time-Series dengan LSTM

Tahap keempat adalah prediksi harga saham menggunakan Long Short-Term Memory (LSTM), yang merupakan model Recurrent Neural Network (RNN) yang dirancang khusus untuk menangani data deret waktu (time-series). Dalam tahap ini, dilakukan pembentukan dataset berurutan menggunakan metode windowing. Model LSTM yang dibangun terdiri dari dua lapisan LSTM serta lapisan Dense sebagai output untuk menghasilkan prediksi harga saham yang lebih akurat. Model ini dilatih menggunakan 80% data training dan 20% data validasi, dengan batch size 16 dan 50 epoch. Setelah pelatihan selesai, model dievaluasi menggunakan RMSE (Root Mean Squared Error) dan MAPE (Mean Absolute Percentage Error) untuk mengukur tingkat kesalahan dalam prediksi harga saham.

3.5. Evaluasi dan Interpretasi Hasil

Pada tahap evaluasi dan interpretasi hasil, yang dilakukan untuk menilai efektivitas metode yang diterapkan dalam penelitian ini. Evaluasi klasterisasi dilakukan dengan menghitung Silhouette Score guna mengukur kualitas cluster yang dihasilkan oleh metode K-Means dan DBSCAN. Sedangkan evaluasi prediksi dilakukan menggunakan MSE (Mean Squared Error), RMSE (Root Mean Squared Error), dan MAPE (Mean Absolute Percentage Error) untuk mengukur sejauh mana perbedaan antara harga prediksi dan harga aktual saham NVIDIA. Selain itu, hasil prediksi dibandingkan dengan data historis melalui visualisasi grafik untuk memahami pola yang terbentuk dan menilai keakuratan model dalam memproyeksikan harga saham. Dengan mengikuti tahapan penelitian ini, penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam mengenai pola pergerakan saham NVIDIA serta menguji efektivitas metode klasterisasi dan prediksi dalam analisis pasar saham.

4. HASIL DAN PEMBAHASAN

Dataset yang didapat melalui yahoo finance dibantu dengan menggunakan program python untuk mendapatkannya diperoleh data sebanyak 1292 data seperti yang terlihat pada tabel 1.

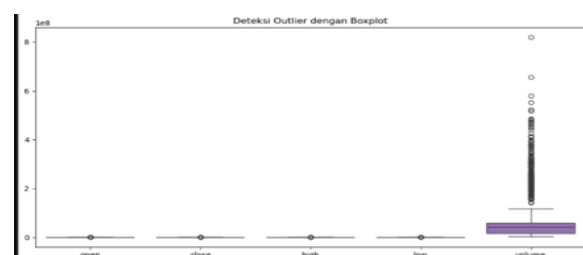
Tabel 1. Dataset saham

No	Date	Open	High	Low	Close	volume
1	1/2/2020	238.75	239.91	236.72	239.91	5941969
2	1/3/2020	235.1	237.83	234.1	236.07	5144308
3	1/6/2020	232.32	237.27	231.27	237.06	6572812
...						
1292	2/21/2025	140.04	141.46	134.03	134.43	225303541

Setelah data dikumpulkan, maka dilakukan tahap preprocessing untuk memastikan kualitas data sebelum digunakan dalam analisis. Tahapan preprocessing ini mencakup beberapa langkah penting, yaitu pengecekan missing values, identifikasi outlier, normalisasi data, serta pembagian dataset menjadi data training dan validasi untuk prediksi nanti. Pengecekan missing values dilakukan untuk mengidentifikasi apakah terdapat data yang hilang, dan jika ditemukan, data tersebut akan diisi menggunakan metode interpolasi atau dihapus jika tidak memungkinkan untuk diisi ulang. Selanjutnya, dilakukan identifikasi outlier menggunakan visualisasi boxplot, yang bertujuan untuk mendeteksi adanya anomali dalam pergerakan harga saham yang dapat mempengaruhi hasil analisis.

```
Jumlah Missing Values per Kolom:
open      0
high      0
low       0
close     0
volume    0
dtype: int64
```

Gambar 2. Pengecekan Missing Values



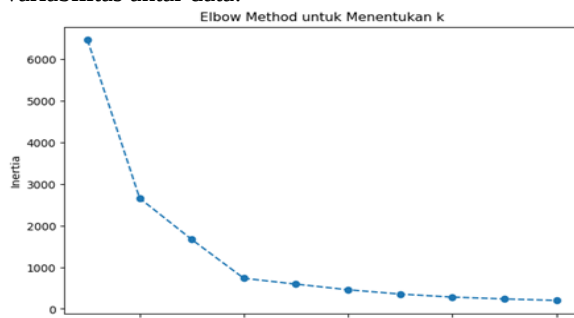
Gambar 3. Identifikasi Outlier

```
Dataset setelah preprocessing:
      open      high      low      close      volume
date
2020-01-02  0.127732  0.118442  0.133665  0.125279  0.004281
2020-01-03  0.124554  0.116637  0.131267  0.121867  0.003304
2020-01-06  0.122133  0.116152  0.128676  0.122747  0.005054
2020-01-07  0.127253  0.120056  0.133363  0.125297  0.006778
2020-01-08  0.128611  0.120291  0.134974  0.125696  0.005492
Preprocessing selesai. Data training dan validasi telah disimpan.
```

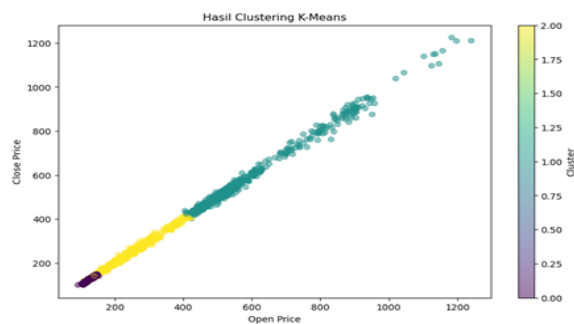
Gambar 4. Normalisasi Data dan Pembagian Data

Setelah dilakukan preprocessing data akan kita klusterisasi menggunakan algoritma K-MEANS Metode Elbow digunakan untuk menentukan jumlah

cluster optimal dalam analisis K-Means. Grafik Elbow pada gambar 5 menunjukkan hubungan antara jumlah cluster dan nilai inertia, yaitu jumlah kuadrat jarak antara setiap titik data dengan centroid cluster terdekatnya. Dari hasil analisis, terlihat bahwa titik siku (elbow point) berada pada $k = 3$, yang menandakan bahwa tiga cluster adalah pilihan optimal. Setelah titik ini, penurunan nilai inertia melambat, menunjukkan bahwa penambahan jumlah cluster tidak lagi memberikan pengurangan signifikan dalam variabilitas antar data.



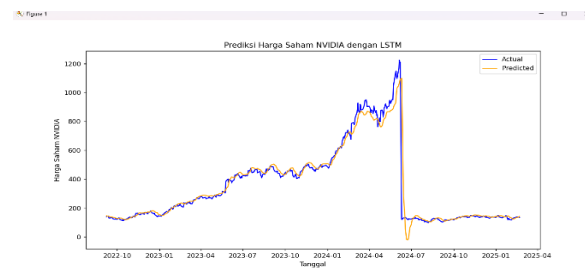
Gambar 5. Hasil Metode ELBOW



Gambar 6. Hasil Klustering

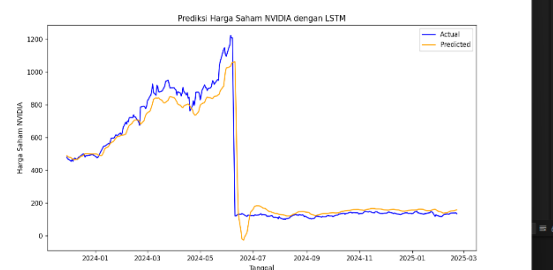
Visualisasi hasil clustering menunjukkan bahwa data telah dikelompokkan menjadi tiga cluster berdasarkan fitur Open Price dan Close Price. Data dengan harga saham yang lebih rendah, sedang, dan tinggi dikelompokkan ke dalam cluster yang berbeda, sebagaimana ditunjukkan oleh warna yang berbeda dalam scatter plot. Hasil dari Silhouette Score yang diperoleh sebesar 0.579 menunjukkan bahwa hasil clustering cukup baik. Nilai ini menunjukkan bahwa sebagian besar data berada dalam cluster yang tepat dengan jarak antar-cluster yang cukup jelas. Skor ini berada dalam kisaran 0.5 - 0.6, yang mengindikasikan struktur cluster yang cukup kuat meskipun ada beberapa titik data yang mungkin berada di batas antar-cluster. Saham NVIDIA cenderung mengalami pola pergerakan harga yang tersegmentasi berdasarkan tingkat harga. Klaster pertama (ungu) mencakup harga saham yang lebih rendah, klaster kedua (kuning) berada di kisaran harga menengah, sedangkan klaster ketiga (hijau kebiruan) mewakili saham dengan harga yang lebih tinggi. Ini menunjukkan bahwa harga saham tidak berubah secara acak, melainkan mengikuti pola tertentu yang dapat diidentifikasi melalui klusterisasi.

Model LSTM digunakan untuk memprediksi harga saham NVIDIA berdasarkan dataset historis. Model dilatih menggunakan jumlah data training yang berbeda (60%, 75%, dan 80% data) dengan 20 epoch. Hasil evaluasi model ditunjukkan melalui nilai RMSE (Root Mean Squared Error) dan MAPE (Mean Absolute Percentage Error).



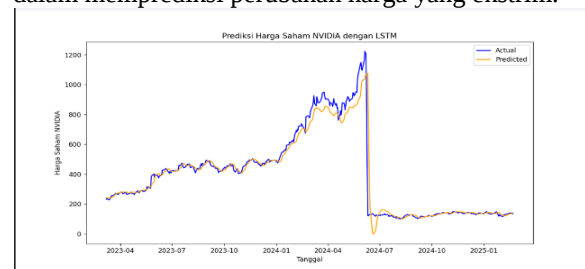
Gambar 7. Grafik Prediksi 60% data train

Pada grafik hasil prediksi gambar 7 didapatkan hasil RMSE: 70.1783 dan MAPE: 9.61%, terlihat juga bahwa model cukup baik dalam mengikuti pola harga saham, tetapi masih terdapat beberapa deviasi antara nilai aktual dan prediksi. Dengan jumlah data training yang lebih sedikit (60 data), model mampu mempelajari tren harga saham dengan baik untuk sebagian besar data, tetapi ketika ada kenaikan harga saham yang tajam, model masih mengalami kesulitan dalam menyesuaikan prediksi.



Gambar 8. Grafik Prediksi 75% data train

Saat jumlah data training ditingkatkan menjadi 75, terlihat bahwa RMSE dan MAPE meningkat menjadi 89.6487 dan 13.94% dibandingkan dengan model sebelumnya. Hal ini mengindikasikan bahwa model masih kesulitan menangkap pola harga saham yang mengalami lonjakan tajam saat validasi. Meskipun lebih banyak data digunakan untuk pelatihan, model tetap mengalami overfitting terhadap data dengan tren yang lebih stabil, sehingga sulit dalam memprediksi perubahan harga yang ekstrim.



Gambar 9. Grafik Prediksi 80% data train

Ketika jumlah data training ditambah menjadi 80, nilai RMSE dan MAPE justru meningkat menjadi 98.7012 dan 21.04%. Hal ini menunjukkan bahwa meskipun lebih banyak data digunakan, model masih mengalami kesulitan dalam memprediksi lonjakan harga saham yang terjadi pada data validasi. Penyebab utama kesalahan prediksi yang besar adalah karena dalam data training tidak terdapat pola harga saham yang mengalami lonjakan drastis, sehingga saat diuji dengan data validasi yang memiliki lonjakan harga signifikan, model kesulitan menyesuaikan hasil prediksinya.

Performa model dipengaruhi oleh beberapa faktor, seperti tidak adanya lonjakan harga di data training yang menyebabkan model kesulitan menangkap pola ekstrem saat diuji dengan data validasi. Selain itu, jumlah epoch yang hanya 20 membuat model masih terbatas dalam mempelajari pola harga saham yang kompleks, sehingga peningkatan epoch mungkin dapat membantu generalisasi yang lebih baik. Kurangnya fitur tambahan, seperti volume perdagangan, berita keuangan, atau sentimen pasar, juga menghambat akurasi prediksi karena model hanya mengandalkan harga saham historis. Terakhir, model dapat mengalami underfitting jika jumlah data training terlalu sedikit atau overfitting jika tidak mampu menangkap pola harga ekstrem dengan baik.

5. KESIMPULAN DAN SARAN

Penelitian ini menganalisis pola dan pergerakan harga saham NVIDIA menggunakan K-Means untuk klusterisasi dan LSTM untuk prediksi. Hasil klusterisasi menunjukkan bahwa jumlah klaster optimal adalah tiga, yang membantu dalam mengidentifikasi pola harga saham berdasarkan kemiripan pergerakan. Namun, metode LSTM dalam prediksi harga saham mengalami kesulitan dalam menangkap pola lonjakan harga yang ekstrem, terutama karena data training tidak memiliki variasi harga yang cukup signifikan. Pengujian dengan berbagai skenario pembagian data (60%-40%, 75%-25%, dan 80%-20%) menunjukkan bahwa semakin besar data training, semakin tinggi kesalahan prediksi, dengan RMSE dan MAPE meningkat pada validasi. Faktor utama yang mempengaruhi hasil ini adalah kurangnya data lonjakan harga dalam training, terbatasnya fitur yang digunakan, serta kemungkinan underfitting atau overfitting pada model. Dengan demikian, K-Means terbukti efektif dalam memahami pola harga saham, tetapi LSTM memiliki keterbatasan dalam menangani lonjakan harga yang tiba-tiba. Untuk meningkatkan performa model, perlu ditambahkan fitur seperti volume perdagangan atau indikator teknikal, serta mempertimbangkan metode lain yang lebih adaptif terhadap volatilitas harga saham.

DAFTAR PUSTAKA

- [1] Hartanti, N. T. (2020). Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional. *J. Nas. Teknol. dan Sist. Inf*, 6(2), 82-89.
- [2] Harani, N. H., Prianto, C., & Nugraha, F. A. (2020). Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python. *Jurnal Manajemen Informatika (JAMIKA)*, 10(2), 133-146.
- [3] Achyar, G., & Rohaeni, O. (2022). Penggunaan Hybrid K-Means dan General Regression Neural Network untuk Prediksi Harga Saham Indeks LQ45. *Jurnal Riset Matematika*, 111-120.
- [4] Chen, Q., Zhang, W., & Lou, Y. (2020). Forecasting stock prices using a hybrid deep learning model integrating attention mechanism, multi-layer perceptron, and bidirectional long-short term memory neural network. *Ieee Access*, 8, 117365-117376.
- [5] Pranata, K. S., Gunawan, A. A., & Gaol, F. L. (2023). Development clustering system IDX company with k-means algorithm and DBSCAN based on fundamental indicator and ESG. *Procedia Computer Science*, 216, 319-327.
- [6] Bachtiar, A., Rizqi, N. I., & Mubarak, S. (2023). Komparasi Metode Clustering K-Means, DbSCAN dan Hierarchical Untuk Analisis Penyakit Hepatitis C. *Journal Information System Development (ISD)*, 8(1), 27-32.
- [7] Saputra, A., & Yusuf, R. (2024). Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM: Comparison of the DBSCAN and K-MEANS Algorithms in Segmenting Customers Using Public Transportation of Transjakarta Using the RFM Method. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1346-1361.
- [8] Riani, A. P., Voutama, A., & Ridwan, T. (2023). Penerapan K-Means clustering dalam pengelompokan hasil belajar peserta didik dengan metode Elbow. *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, 6(1), 164-172.
- [9] Rizkilloh, M. F., & Widiyanesti, S. (2022). Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term Memory (LSTM). *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(1), 25-31.
- [10] Setiawan, L., Susanti, D., & Riaman, R. (2023). Analisis Perbandingan Hasil Peramalan Harga Saham Menggunakan Model Autoregresive Integrated Moving Average dan Long Short Term Memory. *Jurnal Matematika Integratif*, 19(2), 223-234.