

IMPLEMENTASI ARSITEKTUR DATA *PIPELINE SPARK – AIRFLOW* UNTUK PERAMALAN HARGA SAHAM ICBP, MYOR, DAN GOOD MENGGUNAKAN LSTM

Ilil Musyarof Asfiani^{1*}

¹Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran”
Jawa Timur, Surabaya, Indonesia

*Penulis korespondensi: iyilmusyawa2@gmail.com

ABSTRAK

Industri makanan dan minuman merupakan sektor krusial dalam perekonomian Indonesia, namun menghadapi tantangan berupa volatilitas harga saham yang tinggi. Penelitian ini bertujuan membangun sistem prediksi harga saham otomatis untuk emiten PT Indofood CBP Sukses Makmur Tbk (ICBP), PT Mayora Indah Tbk (MYOR), dan PT Garudafood Putra Putri Jaya Tbk (GOOD). Metode yang digunakan adalah algoritma *Long Short-Term Memory* (LSTM) yang diintegrasikan dalam arsitektur data *pipeline* berbasis Big Data. Penelitian ini memanfaatkan *Apache Spark* untuk meningkatkan efisiensi *pre-processing* data historis serta *Apache Airflow* sebagai orkestrator untuk mengotomatisasi seluruh alur kerja, mulai dari akuisisi data melalui *API Yahoo Finance*, penyimpanan terstruktur di PostgreSQL, hingga tahap pelatihan model. Kinerja model dievaluasi menggunakan metrik RMSE, MAE, dan MAPE untuk menilai akurasi serta stabilitas prediksi. Hasil penelitian menunjukkan bahwa implementasi LSTM dalam arsitektur *pipeline* mampu memberikan performa prediksi yang tinggi dan konsisten. Model mencatatkan akurasi tertinggi pada saham GOOD sebesar 98,38% dengan MAPE 1,62%, disusul ICBP (98,08%) dan MYOR (97,57%). Efektivitas *pipeline* juga terlihat dari waktu pemrosesan yang stabil serta kemampuan sistem mempertahankan performa meski dijalankan secara terjadwal melalui *Airflow*.

Kata kunci: Prediksi Harga Saham, *Long Short-Term Memory* (LSTM), *Apache Spark*, *Apache Airflow*, Data Pipeline.

1 PENDAHULUAN

Industri makanan dan minuman merupakan salah satu sektor yang berperan penting dalam perekonomian Indonesia karena kontribusinya terhadap pertumbuhan industri dan stabilitas pasar domestik. Namun, sektor ini juga menghadapi tantangan berupa volatilitas harga saham yang tinggi, yang dipengaruhi oleh dinamika pasar, perubahan kebijakan, serta faktor makroekonomi lainnya. Kondisi tersebut menyebabkan analisis dan prediksi harga saham menjadi kebutuhan penting bagi investor maupun pelaku industri yang memerlukan informasi akurat untuk pengambilan keputusan. Sejumlah penelitian menunjukkan bahwa pendekatan statistik tradisional memiliki keterbatasan dalam memodelkan pola non-linear pada data deret waktu saham, sehingga diperlukan metode yang lebih adaptif dan akurat (Wang, 2019; Lim & Zohren, 2020).

Perkembangan teknologi Big Data memungkinkan pemrosesan data dalam skala besar dan kecepatan tinggi melalui arsitektur komputasi terdistribusi. *Apache Spark* menjadi salah satu teknologi yang banyak digunakan dalam pengolahan data historis karena kemampuannya melakukan transformasi data secara efisien (Gao, 2021). Di sisi lain, orkestrator seperti *Apache Airflow* berperan penting dalam mengotomatisasi alur kerja (*workflow*) sehingga pengumpulan,

pembersihan, dan pemodelan data dapat berjalan secara terjadwal dan konsisten (Gusein-Zade, 2022). Integrasi kedua teknologi ini membuka peluang penerapan *pipeline* data modern untuk mendukung analisis pasar saham secara *real-time* maupun batch.

Dalam ranah pemodelan prediksi, algoritma *Long Short-Term Memory* (LSTM) terbukti unggul dalam menangkap pola sekuensial jangka panjang pada data *time-series*, termasuk harga saham yang memiliki karakteristik *non-stasioner* dan *fluktuatif* (ATLAS Collaboration, 2017). Sejumlah studi melaporkan bahwa LSTM memberikan akurasi yang lebih baik dibandingkan metode konvensional seperti ARIMA dan regresi linier ketika digunakan untuk peramalan finansial (Bińkowski et al., 2018). Dengan demikian, masih terdapat kesenjangan penelitian terkait integrasi model prediksi berbasis LSTM dengan arsitektur data *pipeline* terotomatisasi yang mampu menjamin konsistensi, efisiensi, dan skalabilitas pengolahan data saham. Berdasarkan keterbatasan penelitian sebelumnya yang belum mengintegrasikan model prediksi dengan arsitektur data *pipeline* terotomatisasi, penelitian ini mengajukan hipotesis bahwa integrasi *Apache Spark* dan *Apache Airflow* dalam *pipeline* data dapat meningkatkan akurasi dan stabilitas prediksi harga saham berbasis LSTM dibandingkan pendekatan konvensional.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan mengimplementasikan arsitektur data *pipeline* menggunakan *Apache Spark* dan *Apache Airflow* untuk peramalan harga saham emiten ICBP, MYOR, dan GOOD dengan algoritma LSTM. *Pipeline* dirancang untuk mencakup proses akuisisi data otomatis, *preprocessing* terdistribusi, penyimpanan berlapis, serta pelatihan model secara terjadwal. Penelitian ini diharapkan mampu menghasilkan model prediksi yang lebih akurat, stabil, dan efisien dibandingkan pendekatan LSTM konvensional, sekaligus memberikan gambaran mengenai efektivitas integrasi teknologi Big Data dalam sistem prediksi harga saham modern.

2 METODE

2.1 Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan membangun arsitektur data *pipeline* untuk mendukung proses prediksi harga saham. Alur penelitian terdiri dari beberapa tahap utama, yaitu pengumpulan data historis, penyimpanan data pada beberapa lapisan basis data, *preprocessing* menggunakan *Apache Spark*, pelatihan model *Long Short-Term Memory* (LSTM), evaluasi performa prediksi, serta visualisasi hasil menggunakan *Matplotlib*. Seluruh proses tersebut dijalankan secara otomatis dan terjadwal melalui *Apache Airflow* (Irmayani & Jayanti, 2025).

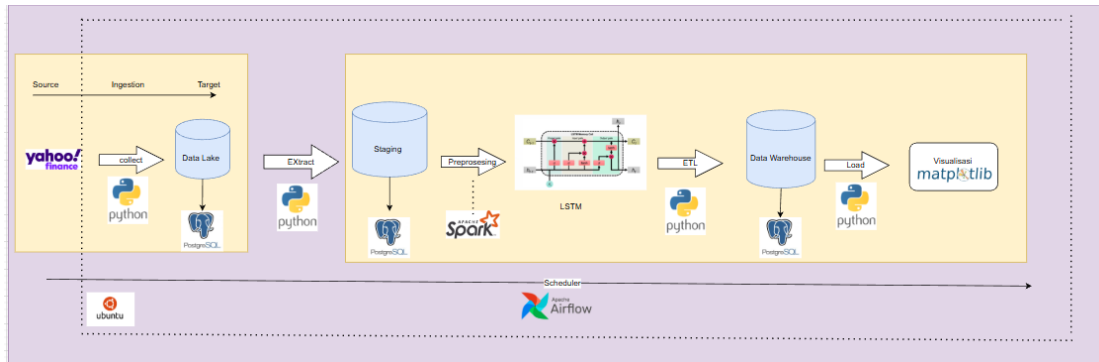
2.1.1 Waktu, Tempat, dan Alat Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan membangun arsitektur data *pipeline* untuk mendukung proses prediksi harga saham. Alur penelitian terdiri dari beberapa tahap utama, yaitu pengumpulan data historis, penyimpanan data pada beberapa lapisan basis data, *preprocessing* menggunakan *Apache Spark*, pelatihan model *Long Short-Term Memory* (LSTM), evaluasi performa prediksi, serta visualisasi hasil menggunakan *Matplotlib*. Seluruh proses tersebut dijalankan secara otomatis dan terjadwal melalui *Apache Airflow* (Karim et al., 2017).

2.1.2 Arsitektur Data Pipeline

Arsitektur sistem prediksi harga saham yang diusulkan pada penelitian ini ditunjukkan pada Gambar 1. Sistem dirancang dalam bentuk data *pipeline* terotomatisasi yang mengintegrasikan proses pengambilan data, penyimpanan berlapis, *preprocessing* terdistribusi, pelatihan model, hingga visualisasi hasil prediksi. Seluruh alur kerja dijalankan pada lingkungan komputasi berbasis

Linux dan diorkestrasi menggunakan *Apache Airflow* untuk memastikan setiap tahapan berjalan secara terjadwal dan sesuai dengan dependensi antar-proses.



Gambar 1. Model Integrasi *AirFlow*

Berdasarkan Gambar 1, data historis harga saham diperoleh dari Yahoo Finance API dan dikumpulkan menggunakan skrip Python, kemudian disimpan ke dalam lapisan Data Lake pada PostgreSQL sebagai data mentah. Selanjutnya, data diekstraksi ke lapisan Staging untuk dilakukan pembersihan awal sebelum diproses lebih lanjut menggunakan Apache Spark pada tahap *preprocessing*. Hasil *preprocessing* digunakan sebagai input pada proses pelatihan model Long Short-Term Memory (LSTM). Model yang telah dilatih kemudian menghasilkan data prediksi yang disimpan pada Data Warehouse dan divisualisasikan menggunakan Matplotlib. Seluruh rangkaian proses tersebut diatur dan dijadwalkan secara otomatis melalui Apache Airflow (Ogeawuchi et al., 2022a).

2.2 Pengumpulan Data

Pengumpulan data dilakukan dengan memanfaatkan *Yahoo Finance API* untuk memperoleh data historis harga saham harian dari tiga emiten sektor makanan dan minuman, yaitu PT Indofood CBP Sukses Makmur Tbk (ICBP), PT Mayora Indah Tbk (MYOR), dan PT Garudafood Putra Putri Jaya Tbk (GOOD). Data yang digunakan mencakup periode Januari 2023 hingga Desember 2025, dengan atribut harga pembukaan (*open*), harga penutupan (*close*), harga tertinggi (*high*), harga terendah (*low*), volume transaksi, serta tanggal perdagangan. Pemilihan rentang waktu tersebut bertujuan untuk menangkap pola pergerakan harga saham terkini serta mencerminkan kondisi pasar yang relevan terhadap dinamika ekonomi terbaru. Proses pengambilan data dilakukan secara otomatis melalui task khusus pada *Apache Airflow* yang dijalankan secara terjadwal (Siami Namin & Siami Namin, 2018).

2.3 Penyimpanan Data

Data hasil pengambilan dari *Yahoo Finance API* disimpan menggunakan pendekatan penyimpanan berlapis untuk menjaga konsistensi dan keterlacakan data. Lapisan pertama adalah *Data Lake*, yang berfungsi sebagai penyimpanan data mentah tanpa transformasi. Selanjutnya, data dipindahkan ke *Data Staging*, yaitu lapisan yang menampung data yang telah melalui pembersihan dasar seperti validasi tipe data, penghapusan duplikasi, dan penanganan nilai hilang (*missing values*). Lapisan terakhir adalah *Prepared Layer*, yang menyimpan data hasil *preprocessing* dan telah siap digunakan sebagai input pada tahap pelatihan model. Pemisahan lapisan penyimpanan ini bertujuan untuk meningkatkan keandalan *pipeline* dan memudahkan reproduksi proses penelitian (Köppe & Zhou, 2016).

2.4 *Preprocessing* Menggunakan Apache Spark

Tahap *preprocessing* dilakukan menggunakan *Apache Spark* untuk mengolah data secara efisien dalam skala besar. Proses *preprocessing* meliputi pengurutan data berdasarkan tanggal perdagangan, normalisasi nilai numerik menggunakan metode *MinMaxScaler*, serta pembuatan fitur *lag* hingga 10 hari untuk menangkap pola historis harga saham. Selain itu, struktur data ditransformasikan ke dalam bentuk sekuensial agar sesuai dengan kebutuhan input model *Long Short-Term Memory* (LSTM). Hasil dari tahap *preprocessing* ini kemudian disimpan pada lapisan *prepared* sebagai dataset final yang digunakan pada proses pelatihan model (Zaharia et al., 2010).

2.5 Pelatihan Model LSTM

Pelatihan model dilakukan menggunakan algoritma *Long Short-Term Memory* (LSTM) untuk mempelajari pola sekuensial pada data harga saham. Dataset dibagi menjadi dua bagian, yaitu 80% data untuk pelatihan dan 20% data untuk pengujian. Model dilatih menggunakan *optimizer Adam* dengan fungsi kerugian (*loss function*) *Mean Squared Error* (MSE). Parameter pelatihan yang digunakan meliputi *batch size* sebesar 32, jumlah *epoch* sekitar 50 *iterasi*, serta panjang *input sequence* sebanyak 10 langkah waktu. Proses pelatihan dijalankan secara otomatis melalui *Apache Airflow*, dan model yang telah dilatih disimpan dalam format *.h5* untuk digunakan pada tahap prediksi (Kingma & Ba, 2017).

2.6 Evaluasi Model

Evaluasi performa model dilakukan untuk menilai tingkat akurasi prediksi harga saham yang dihasilkan. Tiga metrik evaluasi digunakan dalam penelitian ini, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE). Ketiga metrik tersebut digunakan untuk mengukur selisih antara nilai prediksi dan nilai aktual harga saham. Nilai MAE, RMSE, dan MAPE yang lebih kecil menunjukkan bahwa model memiliki performa prediksi yang lebih baik dan stabil (Yi et al., 2021).

2.7 Visualisasi

Hasil prediksi harga saham divisualisasikan menggunakan *Matplotlib* dalam bentuk grafik deret waktu. Visualisasi yang dihasilkan meliputi perbandingan antara harga aktual dan harga hasil prediksi, grafik prediksi jangka pendek untuk tujuh hari ke depan, serta grafik evaluasi performa berdasarkan nilai MAE, RMSE, dan MAPE. Seluruh hasil visualisasi disimpan secara otomatis dalam format file PNG melalui *pipeline Apache Airflow*, sehingga dapat digunakan sebagai bahan analisis dan dokumentasi penelitian (Waskom, 2021).

2.8 Orkestrasi Pipeline dengan Apache Airflow

Apache Airflow digunakan sebagai orkestrator untuk mengatur alur kerja dan ketergantungan antar-tahap dalam *pipeline* prediksi harga saham. Setiap tahapan proses, mulai dari pengambilan data, penyimpanan ke staging, *preprocessing*, pelatihan model, prediksi, hingga visualisasi, direpresentasikan sebagai *task* dalam sebuah DAG. Airflow memastikan setiap *task* dijalankan sesuai urutan yang telah ditentukan, menyediakan mekanisme *logging*, serta mampu melakukan *retry* secara otomatis apabila terjadi kegagalan pada salah satu tahapan proses (Ogeawuchi et al., 2022b).

3 HASIL DAN PEMBAHASAN [TNR 12, CETAK TEBAL, KAPITAL DAN RATA KIRI KANAN]

3.1 Pengumpulan Data

Data penelitian dikumpulkan secara otomatis melalui *Yahoo Finance API* untuk tiga emiten sektor makanan dan minuman, yaitu ICBP, MYOR, dan GOOD. Data historis yang diperoleh mencakup harga pembukaan, penutupan, tertinggi, terendah, volume, serta timestamp perdagangan. Seluruh proses pengambilan data dijalankan secara terjadwal, sehingga pembaruan data dapat dilakukan tanpa intervensi manual.

3.2 Penyimpanan Data Pada Data Lake

Data mentah yang diperoleh dari *API* disimpan terlebih dahulu pada lapisan Data Lake menggunakan *PostgreSQL* tanpa dilakukan transformasi. Penyimpanan ini memastikan seluruh rekaman historis tetap tersedia secara utuh, mendukung proses audit, dan memungkinkan replikasi proses *preprocessing* maupun pelatihan model

3.3 Data Staging

Tahap *Data Staging* dilakukan untuk menyiapkan data sebelum masuk ke proses *preprocessing* Spark. Perbaikan data meliputi validasi tipe data, pemrosesan *missing values*, penghilangan duplikasi, dan penyeragaman format waktu. Hasilnya adalah dataset yang bersih dan konsisten, disimpan dalam skema khusus pada basis data.

3.4 Preprocessing Menggunakan Apache Spark

Apache Spark digunakan sebagai mesin pemrosesan data untuk mengolah dataset berskala besar. Tahapan *preprocessing* mencakup pengurutan data berdasarkan timestamp, normalisasi menggunakan *MinMaxScaler*, pembuatan fitur lag 10 hari menggunakan window function Spark, serta pembentukan struktur tensor 3 dimensi yang diperlukan untuk model LSTM.

3.5 Pelatihan Model LSTM Berbasis Pipeline

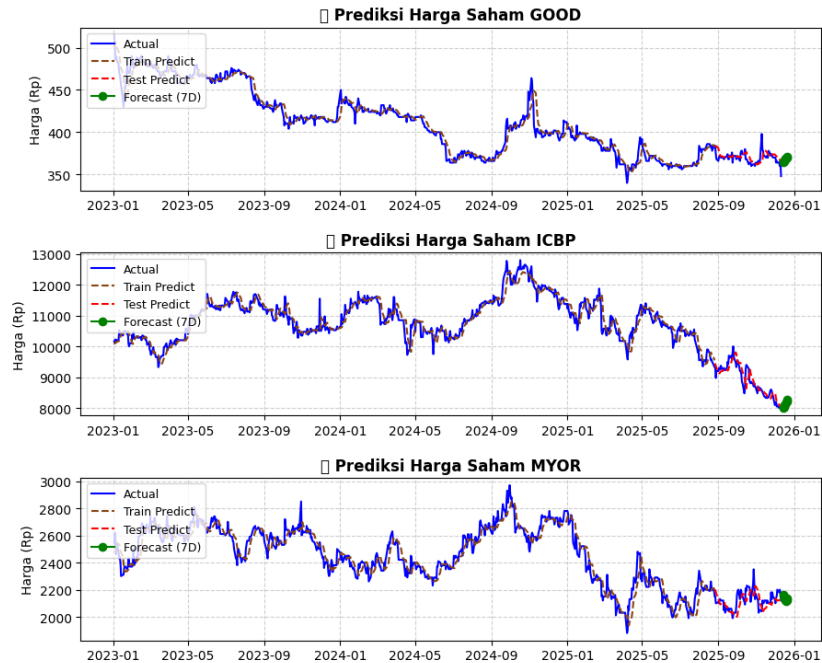
Model *Long Short-Term Memory* (LSTM) dilatih menggunakan dataset yang telah dinormalisasi dan dibentuk menjadi *sequence time-series*. Dataset dibagi menjadi 80% data latih dan 20% data uji. Model dilatih menggunakan *optimizer Adam*, *loss function* MSE, serta jumlah *epoch* sekitar 50 *iterasi* untuk mencapai keseimbangan antara akurasi dan efisiensi. Pelatihan dijalankan otomatis melalui *Airflow* sehingga model dapat diperbarui mengikuti kondisi pasar.

3.6 Penyimpanan Model

Model terbaik dari hasil pelatihan disimpan dalam format H5 untuk mendukung proses prediksi otomatis. Setelah pelatihan, sistem menghitung nilai RMSE, MAE, MAPE, serta akurasi prediksi, kemudian menyimpan hasil evaluasi tersebut ke dalam basis data dengan menghapus catatan lama agar hanya nilai terbaru yang tersimpan.

3.7 Model Prediksi Berbasis Pipeline (Visualisasi Matplotlib)

Model yang telah disimpan digunakan untuk menghasilkan prediksi pada data uji dan prediksi jangka pendek tujuh hari ke depan. Seluruh visualisasi prediksi dibuat menggunakan *Matplotlib* dalam bentuk grafik deret waktu yang menampilkan harga aktual versus prediksi. Plot prediksi diekspor otomatis sebagai file PNG melalui *Airflow* untuk kebutuhan analisis lebih lanjut.



Gambar 2. Visualisasi Prediksi

3.8 Evaluasi Model

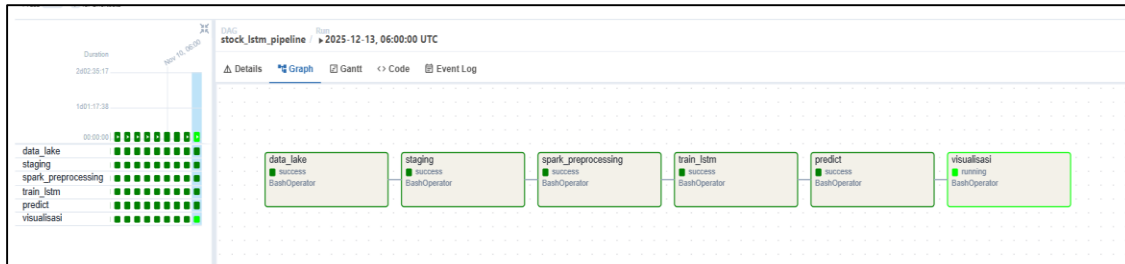
Evaluasi model dilakukan menggunakan metrik RMSE, MAE, dan MAPE. Hasil evaluasi menunjukkan bahwa model *pipeline* mampu memberikan prediksi dengan tingkat akurasi tinggi. Tabel 1 menampilkan hasil evaluasi.

Tabel 1. Hasil Evaluasi Model Prediksi

Saham	RMSE	MAE	MAPE (%)	Akurasi (%)
GOOD	6.81	4.50	1.22	98.78
ICBP	238.13	181.60	1.86	98.14
MYOR	76.63	58.75	2.70	97.30

3.9 Airflow sebagai Orkestrator Pipeline

Apache Airflow digunakan sebagai pengatur alur kerja dalam *pipeline* prediksi harga saham. Setiap tahapan mulai dari pengambilan data, *staging*, *preprocessing*, pelatihan model, hingga prediksi diatur dalam bentuk DAG (*Directed Acyclic Graph*). Airflow memastikan setiap task berjalan sesuai urutan yang benar, menyediakan fitur monitoring, logging, serta penjadwalan otomatis sehingga sistem prediksi dapat berjalan secara stabil dan berkelanjutan.



Gambar 3. Arsitektur *AirFlow*

4 KESIMPULAN

Penelitian ini berhasil mengembangkan sistem prediksi harga saham berbasis arsitektur data *pipeline* yang mengintegrasikan *Apache Spark*, *Apache Airflow*, serta model *Long Short-Term Memory* (LSTM). *Pipeline* yang dibangun mampu mengotomatisasi seluruh proses alur data mulai dari pengumpulan, pembersihan, *preprocessing*, pelatihan model, hingga prediksi harian. Pemanfaatan *Spark* terbukti meningkatkan efisiensi dalam penanganan data berskala besar melalui pembuatan fitur *lag* dan normalisasi yang konsisten, sementara *Airflow* memastikan setiap proses berjalan sesuai urutan dan terjadwal secara berkelanjutan. Hasil evaluasi menunjukkan bahwa model LSTM berbasis *pipeline* memberikan performa yang sangat baik, dengan nilai MAPE di bawah 3% untuk seluruh emiten yang diuji. Saham GOOD mencatat akurasi tertinggi (98,78%), diikuti ICBP (98,14%) dan MYOR (97,30%). Hal ini menunjukkan bahwa model mampu menangkap pola pergerakan harga secara efektif dan memberikan prediksi yang stabil. Visualisasi menggunakan Matplotlib turut memperjelas keselarasan antara harga aktual dan prediksi, baik pada data uji maupun prediksi 7 hari ke depan. Secara keseluruhan, implementasi arsitektur *pipeline Spark–Airflow* terbukti efektif dalam mendukung sistem prediksi harga saham yang otomatis, efisien, dan akurat, sehingga relevan untuk diterapkan pada kasus analitik pasar keuangan berskala lebih luas.

5 SARAN

Penelitian selanjutnya dapat menambahkan variabel eksternal seperti sentimen berita atau indikator makroekonomi untuk meningkatkan akurasi model. Serta penggunaan model dapat dikembangkan menggunakan arsitektur lain seperti GRU atau model lainnya untuk membandingkan performa prediksi. *Pipeline* dapat diperluas agar mendukung data dengan frekuensi lebih tinggi, seperti *data intraday*, sehingga sistem prediksi lebih adaptif terhadap perubahan pasar.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam pelaksanaan penelitian ini, baik dalam bentuk arahan, bantuan teknis, maupun dukungan nonteknis. Ucapan terima kasih juga disampaikan kepada pihak penyedia data yang telah memberikan akses terhadap data yang digunakan dalam penelitian ini sehingga penelitian dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

- ATLAS Collaboration. (2017). *Evidence for the $H \rightarrow b\bar{b}$ decay with the ATLAS detector*. [https://doi.org/10.1007/JHEP12\(2017\)024](https://doi.org/10.1007/JHEP12(2017)024)
- Bińkowski, M., Marti, G., & Donnat, P. (2018). *Autoregressive Convolutional Neural Networks for Asynchronous Time Series*. <http://arxiv.org/abs/1703.04122>
- Gao, X. (2021). *Estimation and Selection Properties of the LAD Fused Lasso Signal Approximator*. <http://arxiv.org/abs/2105.00045>
- Gusein-Zade, S. M. (2022). *Generating series of classes of exotic un-ordered configuration spaces*. <http://arxiv.org/abs/2203.10798>
- Irmayani, W., & Jayanti, W. E. (2025). Visualisasi Prediksi Harga Saham Menggunakan Model Long Short-Term Memory (LSTM). In *Jurnal Sistem Informasi Akuntansi* (Vol. 6, Issue 1). <http://jurnal.bsi.ac.id/index.php/justian>
- Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2017). *LSTM Fully Convolutional Networks for Time Series Classification*. <https://doi.org/10.1109/ACCESS.2017.2779939>
- Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization*. <http://arxiv.org/abs/1412.6980>
- Köppe, M., & Zhou, Y. (2016). *New computer-based search strategies for extreme functions of the Gomory--Johnson infinite group problem*. <https://doi.org/10.1007/s12532-016-0115-9>
- Lim, B., & Zohren, S. (2020). *Time Series Forecasting With Deep Learning: A Survey*. <https://doi.org/10.1098/rsta.2020.0209>
- Ogeawuchi, J. C., Uzoka, A. C., Alozie, C. E., Agboola, O. A., Gbenle, T. P., & Owoade, S. (2022a). Systematic Review of Data Orchestration and Workflow Automation in Modern Data Engineering for Scalable Business Intelligence. *International Journal of Social Science Exceptional Research*, 1(1), 283–290. <https://doi.org/10.54660/ijsser.2022.1.1.283-290>
- Ogeawuchi, J. C., Uzoka, A. C., Alozie, C. E., Agboola, O. A., Gbenle, T. P., & Owoade, S. (2022b). Systematic Review of Data Orchestration and Workflow Automation in Modern Data Engineering for Scalable Business Intelligence. *International Journal of Social Science Exceptional Research*, 1(1), 283–290. <https://doi.org/10.54660/ijsser.2022.1.1.283-290>
- Siami Namin, S., & Siami Namin, A. (2018). *FORECASTING ECONOMIC AND FINANCIAL TIME SERIES: ARIMA VS. LSTM*.
- Wang, J. (2019). *Instability caused by the fermion-fermion interactions combined with rotational and particle-hole asymmetries in three-dimensional materials with quadratic band touching*. <http://arxiv.org/abs/1906.01412>
- Waskom, M. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. <https://doi.org/10.21105/joss.03021>
- Yi, G., Zhou, C., Cao, Y., & Hu, H. (2021). Hybrid assembly path planning for complex products by reusing a priori data. *Mathematics*, 9(4), 1–13. <https://doi.org/10.3390/math9040395>
- Zaharia, M., Chowdhury, M., Franklin, M. J., & Shenker, S. (2010). *Spark: Cluster Computing with Working Sets*.