SKRIPSI

PREDIKSI HARGA SAHAM MAKANAN DAN MINUMAN MENGGUNAKAN *STACKED* LSTM BERBASIS DATA *PIPELINE* DENGAN *APACHE SPARK* DAN *APACHE AIRFLOW*

ILIL MUSYAROF ASFIANI
NPM 21083010073

DOSEN PEMBIMBING
Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU., ASEAN Eng.
Trimono, S.Si., M.Si

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026

LEMBAR PENGESAHAN

PREDIKSI HARGA SAHAM MAKANAN DAN MINUMAN MENGGUNAKAN *STACKED LSTM* BERBASIS DATA *PIPELINE* DENGAN *APACHE SPARK* DAN *APACHE AIRFLOW*

Oleh:
ILIL MUSYAROF ASFIANI
NPM. 21083010073

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada Tanggal 27 April 2026:

Menyetujui,

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU., ASEAN Eng. NIP. 19801205 200501 1 002	....	**(Pembimbing I)**
Trimono, S.Si., M.Si. NIP. 19950908 202203 1 003	....	**(Pembimbing II)**
Aviolla Terza Damaliana, S.Si., M.Stat. NIP. 19940802 202203 2 015	....	**(Ketua Penguji)**
Alfan Rizaldy Pratama, S.Tr.T., M.Tr.Kom. NIP. 19990606 202406 1 001	....	**(Penguji I)**

Mengetahui,

Dekan Fakultas Ilmu Komputer

Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

PREDIKSI HARGA SAHAM MAKANAN DAN MINUMAN MENGGUNAKAN *STACKED LSTM* BERBASIS DATA *PIPELINE* DENGAN *APACHE SPARK* DAN *APACHE AIRFLOW*

Oleh:
ILIL MUSYAROF ASFIANI
NPM. 21083010073

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., ASEAN Eng.
NIP. 19801205 200501 1 002

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa	:	Ilil Musyarof Asfiani
NPM	:	21083010073
Program	:	Sarjana (S1)
Program Studi	:	Sains Data
Fakultas	:	Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 27 April 2026
Yang Membuat Pernyataan,

ILIL MUSYAROF ASFIANI
NPM. 21083010073

ABSTRAK

Nama Mahasiswa / NPM : Ilil Musyarof Asfiani / 21083010073
Judul Skripsi : Prediksi Harga Saham Makanan dan Minuman Menggunakan *Stacked LSTM* Berbasis Data *Pipeline* Dengan *Apache Spark* Dan *Apache Airflow*
Dosen Pembimbing : 1. Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU., ASEAN Eng.
2. Trimono, S.Si., M.Si

Pergerakan harga saham sektor makanan dan minuman memiliki karakteristik *time series* yang fluktuatif, nonlinier, serta dipengaruhi pola (*pattern*) historis, sehingga sulit diprediksi dengan metode konvensional. Selain itu, pengolahan data dalam jumlah besar membutuhkan sistem yang terintegrasi dan otomatis. Penelitian ini bertujuan untuk merancang sistem prediksi harga saham menggunakan model *Stacked Long Short-Term Memory* (LSTM) yang terintegrasi dengan *data pipeline* berbasis *Apache Spark* dan *Apache Airflow*. Metode penelitian meliputi pembangunan *data pipeline* (data *lake, staging, preprocessing*), serta pemodelan *three-layer stacked* LSTM. Pengujian dilakukan menggunakan empat *pattern* konfigurasi unit LSTM (128-64-32, 64-64-32, 128-128-32, dan 200-200-200). Analisis *time series* menggunakan uji ADF dan ACF-PACF menunjukkan data menjadi stasioner setelah *differencing* (p-value 0,0000). Pola temporal ditangkap melalui fitur lag dengan lag optimal 1 (GOOD), 4 (ICBP), dan 3 (MYOR), serta *global lag* sebesar 4. Hasil menunjukkan model memiliki performa yang baik dengan tingkat kesalahan rendah. Performa terbaik diperoleh pada saham MYOR (*pattern* 2) dengan RMSE 65,50, MAE 50,47, dan MAPE 2,28%; saham ICBP (*pattern* 2) dengan RMSE 277,48, MAE 218,84, dan MAPE 2,36%; serta saham GOOD (*pattern* 3) dengan RMSE 11,68, MAE 9,51, dan MAPE 2,57%. Secara keseluruhan, nilai MAPE berada pada kisaran 2%–4% yang menunjukkan akurasi model cukup baik. Selain itu, *data pipeline* berjalan otomatis dengan keberhasilan *task* 100% dan durasi proses yang efisien. Kesimpulannya, model *Stacked* LSTM dengan konfigurasi *pattern* yang tepat mampu menangkap pola temporal secara efektif dan menghasilkan sistem prediksi yang akurat, efisien, serta terotomatisasi.

Kata Kunci: *Apache Airflow*, *Apache Spark*, *Data Pipeline,* Prediksi Harga Saham*, Three-Layer Stacked* LSTM

ABSTRACT

Student Name / NPM	:	Ilil Musyarof Asfiani/ 21083010073
Undergraduate thesis title	:	*Food and Beverage Stock Price Prediction Using Stacked LSTM Based on Data Pipeline with Apache Spark and Apache Airflow*
Advisors	:	1. Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU., ASEAN Eng. *2.* Trimono, S.Si., M.Si

Stock price movements in the food and beverage sector exhibit time series characteristics that are fluctuative, nonlinear, and influenced by historical patterns, making them difficult to predict using conventional methods. In addition, processing large-scale stock data requires an integrated and automated system. This study aims to design a stock price prediction system using the Stacked Long Short-Term Memory (LSTM) model integrated with a data pipeline based on Apache Spark and Apache Airflow. The research method includes the development of a data pipeline (data lake, staging, preprocessing) and the implementation of a three-layer stacked LSTM model. Model evaluation was conducted using four LSTM unit configuration patterns (128-64-32, 64-64-32, 128-128-32, and 200-200-200). Time series analysis using ADF and ACF-PACF tests indicates that the data becomes stationary after differencing (p-value 0.0000). Temporal patterns are captured through lag features with optimal lag values of 1 (GOOD), 4 (ICBP), and 3 (MYOR), and a global lag of 4. The results show that the model achieves good performance with relatively low error rates. The best performance for MYOR stock is obtained using pattern 2 with RMSE of 65.50, MAE of 50.47, and MAPE of 2.28%. For ICBP stock, the best performance is also achieved using pattern 2 with RMSE of 277.48, MAE of 218.84, and MAPE of 2.36%. Meanwhile, for GOOD stock, the best performance is achieved using pattern 3 with RMSE of 11.68, MAE of 9.51, and MAPE of 2.57%. Overall, the MAPE values range from 2% to 4%, indicating that the model has good prediction accuracy. In addition, the data pipeline runs automatically with a 100% task success rate and efficient processing time.In conclusion, the Stacked LSTM model with appropriate pattern configurations is capable of effectively capturing temporal patterns and producing an accurate, efficient, and automated stock price prediction system.

Keywords: *Apache Airflow, Apache Spark, Data Pipeline, Long Short-Term Memory (LSTM), Stock Price Prediction*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

DAFTAR NOTASI

C_t	:	*Cell state* (keadaan sel) pada waktu t
C_{t-1}	:	*Cell state* sebelumnya (pada waktu t -1)
f_t	:	*Forget gate* pada waktu t
i_t	:	*Input gate* pada waktu t
O_t	:	*Output gate* pada waktu t
σ	:	Fungsi aktivasi *sigmoid*
W_{hf} , W_{xf}	:	Bobot untuk *forget gate*
W_{hi} , W_{xi}	:	Bobot untuk *input gate*
W_{ho} , W_{xo}	:	Bobot untuk *output gate*
h_{t-1}	:	*Hidden state* sebelumnya
X_t	:	*Input pada* waktu t
b_f, b_i, b_o	:	Bias untuk *forget fate*, *input gate* dan *output gate*