

## BAB V

### PENUTUP

#### 5.1. Kesimpulan

Berdasarkan hasil perancangan, implementasi, dan pengujian sistem prediksi harga saham sektor makanan dan minuman menggunakan data *pipeline* berbasis *Apache Airflow*, *Apache Spark*, dan model *Stacked Long Short-Term Memory* (LSTM), dapat ditarik kesimpulan sebagai berikut:

1. Perancangan dan implementasi arsitektur data *pipeline* berbasis *Apache Airflow* dan *Apache Spark* dalam pengolahan data saham sektor makanan dan minuman. Arsitektur data *pipeline* berhasil dirancang dan diimplementasikan secara terintegrasi dengan menggunakan *Apache Airflow* sebagai orkestrator alur kerja (*workflow orchestration*) dan *Apache Spark* sebagai mesin pemrosesan data. *Pipeline* dibangun menggunakan pendekatan *Directed Acyclic Graph* (DAG) pada *Airflow* dengan identitas *pipeline stock\_lstm\_pipeline* yang dikonfigurasi untuk berjalan secara otomatis setiap pukul 09.00 WIB menggunakan ekspresi cron (0 9 \* \* \*). *Pipeline* terdiri dari enam *task* yang saling terhubung secara berurutan, yaitu: (1) *data\_lake* untuk akuisisi data historis saham tiga emiten sektor makanan dan minuman, yaitu ICBP.JK, MYOR.JK, dan GOOD.JK dari Yahoo Finance dan penyimpanannya ke dalam skema *data\_lake* pada *database PostgreSQL*; (2) *staging* untuk proses pembersihan dan transformasi data mentah meliputi penghapusan duplikasi, penanganan *missing value* dengan metode *forward fill*, konversi tipe data, dan penyaringan data tidak valid; (3) *spark\_preprocessing* untuk pembentukan fitur *lag* dan variabel target menggunakan *Apache Spark* melalui *window function* berbasis partisi *ticker* dan urutan tanggal; (4) *train\_lstm* untuk pelatihan model; (5) *predict* untuk proses prediksi; serta (6) visualisasi untuk pembuatan grafik hasil prediksi. Setiap *task* dikonfigurasi dengan *trigger rule all\_success* sehingga memastikan eksekusi berjalan secara berurutan sesuai dependensi yang telah ditetapkan. Integrasi *Apache Spark* dalam *pipeline* memungkinkan proses *feature engineering* berbasis *window*

*function* untuk pembentukan fitur *lag* (*close\_lag\_1* hingga *close\_lag\_4*) dan variabel target (*target\_close*) dilakukan secara efisien. Seluruh data dikelola dalam skema bertingkat pada *PostgreSQL*, mulai dari *data\_lake*, *staging*, *prepared*, hingga *predicted*, yang memungkinkan *traceability* dan pemisahan tanggung jawab antartahapan *pipeline* secara jelas.

2. Model *Stacked Long Short-Term Memory* (LSTM) berhasil diintegrasikan ke dalam data *pipeline* secara terstruktur menggunakan Apache Airflow. Integrasi model dilakukan melalui beberapa *task* utama, yaitu *spark\_preprocessing*, *train\_lstm*, dan *predict* yang saling terhubung dalam alur *pipeline*. Pada tahap *spark\_preprocessing*, Apache Spark digunakan untuk melakukan pembentukan fitur *lag* dan variabel target sebagai input model. Selanjutnya, pada *task train\_lstm*, model *Stacked LSTM* dilatih menggunakan data historis saham dengan beberapa kombinasi *pattern* untuk memperoleh konfigurasi terbaik. Model yang telah dilatih kemudian disimpan dalam format *.h5* beserta file *scaler* untuk digunakan kembali pada proses prediksi. Proses prediksi dilakukan pada *task predict* menggunakan metode *autoregressive multi-step forecasting* untuk menghasilkan prediksi harga saham selama 7 hari ke depan. Integrasi model ke dalam *pipeline* memungkinkan seluruh proses mulai dari preprocessing, pelatihan model, hingga prediksi berjalan secara otomatis, terjadwal, dan saling terhubung dalam satu sistem prediksi harga saham yang terintegrasi.
3. Hasil implementasi data *pipeline* dalam menjalankan proses prediksi secara otomatis ditinjau dari keberhasilan eksekusi *task* dan durasi proses pada *Airflow*. Seluruh *task* dalam *pipeline stock\_lstm\_pipeline* berhasil dieksekusi dengan status *success* pada *Apache Airflow* tanpa ditemukan kegagalan maupun kendala selama proses berlangsung. Ditinjau dari keberhasilan eksekusi *task*, setiap *task* berjalan sesuai urutan dependensi yang telah dikonfigurasi melalui *trigger rule all\_success*, sehingga kegagalan pada satu *task* akan menghentikan eksekusi *task* berikutnya dan menjaga integritas alur *pipeline*. Ditinjau dari durasi proses, *task data\_lake* berhasil diselesaikan dalam durasi sekitar 1 detik, yang mencerminkan efisiensi proses pengambilan data

dari Yahoo Finance dan penyimpanannya ke *PostgreSQL*. *Task staging* juga berjalan dalam durasi sekitar 1 detik, menunjukkan bahwa proses pembersihan dan transformasi data berlangsung sangat cepat. *Task spark\_preprocessing* yang melibatkan pembentukan fitur *lag* menggunakan *Apache Spark* berjalan dalam durasi yang singkat dan konsisten, yaitu sekitar 10 detik, yang menunjukkan efisiensi pemrosesan data berbasis *Spark* melalui *window function*. *Task train\_lstm* memerlukan durasi paling lama, yaitu sekitar 4 menit 34 detik, yang mencerminkan kompleksitas komputasi pada proses pelatihan model *Stacked LSTM* dengan mekanisme *grid search* terhadap empat konfigurasi *pattern*. *Task predict* berhasil diselesaikan dalam waktu sekitar 8 detik, menunjukkan bahwa proses prediksi *autoregressive* untuk 7 hari ke depan berjalan secara efisien. *Task visualisasi* memerlukan durasi sekitar 40 detik untuk menghasilkan grafik perbandingan data aktual, prediksi *training*, prediksi *testing*, dan *forecast* ke depan untuk ketiga saham. Secara keseluruhan, *pipeline* mampu menyelesaikan seluruh proses secara otomatis mulai dari akuisisi data hingga visualisasi dalam satu siklus eksekusi terjadwal yang berjalan setiap hari pukul 09.00 WIB.

4. Hasil evaluasi menunjukkan bahwa model *Stacked Long Short-Term Memory* (LSTM) mampu menghasilkan performa prediksi yang baik pada saham sektor makanan dan minuman. Evaluasi dilakukan menggunakan metrik *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE) untuk mengukur tingkat kesalahan prediksi model pada setiap *pattern* yang diuji. Berdasarkan hasil pengujian pada saham MYOR, *pattern* terbaik diperoleh pada *Pattern 2* dengan nilai RMSE sebesar 65,50, MAE sebesar 50,47, dan MAPE sebesar 2,28%, karena menghasilkan nilai error paling rendah dibandingkan *pattern* lainnya. Pada saham ICBP, performa terbaik juga diperoleh pada *Pattern 2* dengan nilai RMSE sebesar 277,48, MAE sebesar 218,84, dan MAPE sebesar 2,36%. Sementara itu, pada saham GOOD, *pattern* terbaik diperoleh pada *Pattern 1* dengan nilai RMSE sebesar 11,77, MAE sebesar 9,34, dan MAPE sebesar 2,49%, meskipun *Pattern 3* memiliki nilai RMSE sedikit lebih rendah, namun *Pattern 1*

menghasilkan kombinasi error yang lebih stabil secara keseluruhan. Secara umum, nilai MAPE pada seluruh saham berada pada kisaran 2% hingga 4%, yang menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang rendah dan performa yang cukup baik dalam memprediksi harga saham. Hasil evaluasi juga menunjukkan bahwa perubahan konfigurasi model pada setiap pattern memberikan pengaruh terhadap performa prediksi, sehingga pemilihan konfigurasi model yang tepat menjadi faktor penting dalam meningkatkan akurasi hasil prediksi.

5. Perancangan dan implementasi antarmuka (GUI) berbasis web menggunakan *Streamlit* untuk visualisasi hasil prediksi harga saham. Antarmuka pengguna (GUI) berbasis web berhasil dirancang dan diimplementasikan menggunakan *framework Streamlit* sebagai lapisan presentasi (*presentation layer*) sistem. *Dashboard* ini terhubung secara langsung dengan *database PostgreSQL* sehingga seluruh data yang ditampilkan merupakan data terkini hasil eksekusi *pipeline*. *Dashboard* terdiri dari tiga halaman utama. Halaman Beranda berfungsi sebagai pusat *monitoring* sistem yang menampilkan informasi arsitektur *pipeline* secara *end-to-end*, ringkasan *dataset*, termasuk jumlah *ticker*, rentang tanggal historis, dan total *record*, serta tabel performa model yang memuat nilai RMSE, MAPE, dan konfigurasi model terbaik untuk setiap *ticker* saham. Halaman *Dashboard* Prediksi merupakan fitur utama sistem yang memungkinkan pengguna memilih *ticker* saham secara interaktif, kemudian sistem secara otomatis memuat dan menampilkan analisis deskriptif data historis berupa statistik harga penutupan dan distribusi harga, parameter model yang digunakan seperti *lag time* dan konfigurasi *dropout*, visualisasi grafik *time series* interaktif yang membandingkan harga aktual dengan prediksi *in-sample*, prediksi *out-sample*, serta *forecast* 7 hari ke depan, serta tabel evaluasi model dan tabel hasil *forecast* numerik. Pengguna dapat mengatur rentang tampilan grafik dan memilih komponen prediksi yang ingin ditampilkan. Halaman Panduan Penggunaan menyediakan dokumentasi sistem yang menjelaskan tujuan aplikasi, fitur setiap halaman, parameter teknis pemodelan, dan panduan *troubleshooting* untuk membantu pengguna dalam mengatasi

kendala seperti kegagalan pemuatan model, masalah koneksi *database*, maupun hasil *forecast* yang belum tersedia. Secara keseluruhan, implementasi *dashboard Streamlit* berhasil menyajikan hasil prediksi model *Stacked LSTM* secara interaktif, informatif, dan mudah digunakan, sehingga mendukung proses analisis dan pengambilan keputusan berdasarkan prediksi harga saham sektor makanan dan minuman.

## 5.2. Saran Pengembangan

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk pengembangan lebih lanjut, antara lain:

1. Penelitian ini menggunakan data harga saham harian dari tiga emiten sektor makanan dan minuman. Untuk meningkatkan representativitas model, penelitian selanjutnya dapat memperluas cakupan data dengan menambah jumlah emiten yang diamati serta mempertimbangkan penggunaan data *intraday* (per jam atau per menit) untuk mendukung analisis prediksi jangka sangat pendek.
2. Beberapa model pada penelitian ini menunjukkan indikasi potensi *overfitting*, khususnya pada saham ICBP dan GOOD. Oleh karena itu, penelitian selanjutnya dapat mempertimbangkan penerapan teknik regularisasi yang lebih komprehensif, seperti *early stopping*, *L1/L2 regularization*, serta *hyperparameter tuning* yang lebih menyeluruh menggunakan metode seperti *Bayesian Optimization* atau *Optuna* untuk menemukan konfigurasi model yang lebih optimal.
3. Penelitian ini hanya menggunakan fitur harga penutupan (*close*) sebagai dasar pembentukan fitur *lag*. Pengembangan lebih lanjut dapat mempertimbangkan penambahan fitur eksternal seperti data sentimen berita keuangan, indikator teknikal (RSI, MACD, *Bollinger Bands*), serta data makroekonomi seperti nilai tukar dan suku bunga untuk meningkatkan kemampuan prediksi model.
4. Untuk meningkatkan skalabilitas sistem pada volume data yang lebih besar, *pipeline* dapat dikembangkan dengan memanfaatkan kapabilitas *distributed computing Apache Spark* secara lebih optimal, misalnya dengan menggunakan

*cluster Spark multi-node* atau mengintegrasikan *message broker* seperti *Apache Kafka* untuk mendukung skenario pemrosesan data secara *real-time (streaming)*.

5. Penelitian mendatang juga dapat mengeksplorasi perbandingan performa model *Stacked LSTM* dengan arsitektur *deep learning* lainnya seperti *Gated Recurrent Unit (GRU)*, *Transformer*, atau model hibrida *CNN-LSTM*, serta membandingkannya dengan model statistik klasik seperti *ARIMA* dan *SARIMA* untuk mendapatkan gambaran yang lebih komprehensif mengenai kemampuan masing-masing metode dalam prediksi harga saham.