

Analisis Perbandingan Model SARIMA dan LSTM dalam Peramalan Produksi Batu Bara di Amerika Serikat

Kelompok - 22

Analisis Data Statistik - A



Meet Our Team



M. Iqbal Nurrifki

5003221061



M. Rifqy Rezvany A.

5003221022

Latar Belakang



**Fluktuasi Produksi
Batu Bara AS yang
Signifikan Setiap
Tahun**



**Peran Strategis Batu
Bara dalam
Ketahanan Energi
Nasional**



**Kebutuhan
Metode Prediksi
yang Akurat dan
Adaptif**

Pendahuluan

Rumusan Masalah

1. Seberapa baik model LSTM dapat memprediksi produksi batu bara di AS dibandingkan dengan model ARIMA berdasarkan akurasi statistik seperti RMSE dan MAPE?
2. Apa kelebihan dan keterbatasan dari masing-masing model dalam konteks data produksi batu bara yang bersifat non-linier dan kompleks?
3. Model manakah yang lebih optimal untuk digunakan dalam konteks peramalan produksi batu bara di Amerika Serikat berdasarkan hasil evaluasi kinerja?

Tujuan

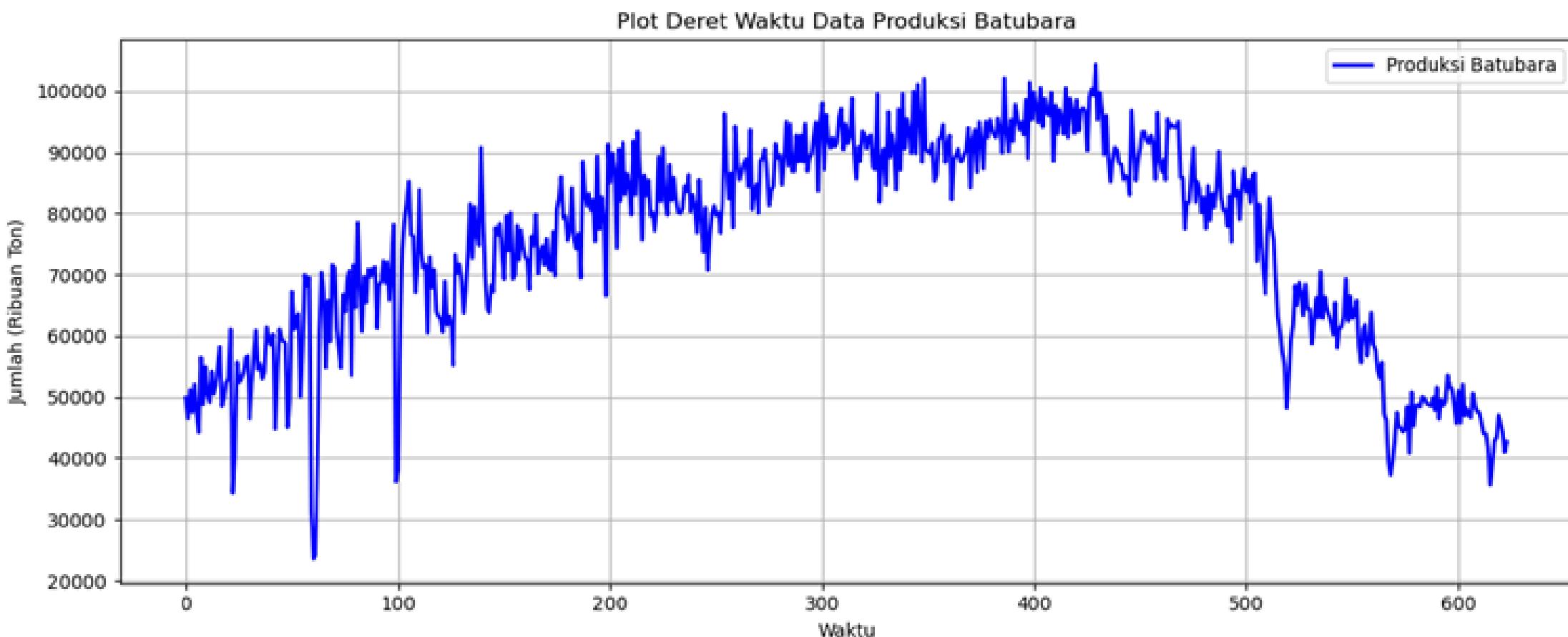
1. Menganalisis kemampuan prediksi LSTM dalam meramalkan produksi batu bara di Amerika Serikat dan membandingkannya dengan model ARIMA berdasarkan akurasi statistik, yaitu nilai RMSE dan MAPE
2. Mengidentifikasi kelebihan dan keterbatasan dari kedua model, ARIMA dan LSTM, dalam menghadapi karakteristik data produksi batu bara yang bersifat non-linier dan kompleks
3. Menentukan model peramalan yang paling optimal berdasarkan hasil evaluasi performa, serta memberikan rekomendasi model terbaik untuk digunakan dalam analisis dan perumusan kebijakan di sektor energi.

Sumber Data

Penelitian ini menggunakan **data sekunder** dari situs resmi U.S. Energy Information Administration (EIA). Data berupa data deret waktu bulanan yang mencakup **produksi batu bara**.

dimana:

- **Periode data:** Januari 1973 – Desember 2024
- **Satuan data:** Ribuan ton



Pengujian Non-Linieritas

Dilakukan uji Terasvirta untuk mendeteksi apakah data produksi batu bara memiliki pola nonlinear. Hasil uji digunakan sebagai dasar dalam memilih model peramalan yang sesuai.

Hipotesis

H_0 : *Model linier cukup untuk menjelaskan pola dalam data.*

H_1 : *Model linier tidak cukup, dan terdapat pola nonlinier dalam data*

Statistik Uji

Chi-square	df	P-value
18,198	2	0,0001118

Keputusan & Kesimpulan

didapatkan keputusan **Tolak H_0** , karena **P-value < 0.05** dan didapat kesimpulan bahwa **data memiliki pola nonlinear** yang signifikan secara statistik. sehingga penggunaan model nonlinear seperti LSTM menjadi lebih relevan dalam konteks peramalan data

Splitting Data

- Data dibagi menjadi **dua bagian**, yaitu **90% untuk pelatihan** (training set) dan **10% untuk pengujian** (testing set). Pembagian ini dilakukan secara berurutan mengikuti struktur deret waktu, tanpa pengacakan, agar mempertahankan urutan temporal data.
- Pemilihan rasio **90:10** dilakukan untuk menyisakan **lima tahun terakhir** sebagai data pengujian guna menangkap pola musiman dan tren terkini, sementara 90% sisanya digunakan untuk pelatihan agar model dapat mempelajari pola historis secara optimal.

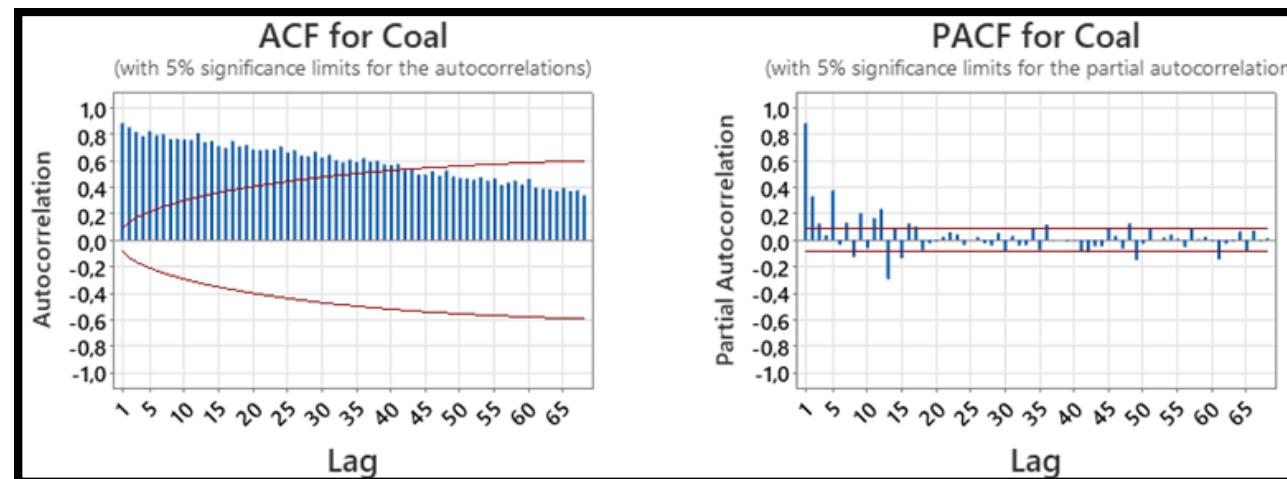
HASIL & PEMBAHASAN

Hasil dan Pembahasan

Prediksi dengan Model SARIMA

Langkah awal yang dilakukan peneliti adalah memvisualisasikan **Autocorrelation Function (ACF)** dan **Partial Autocorrelation Function (PACF)** menggunakan data training.

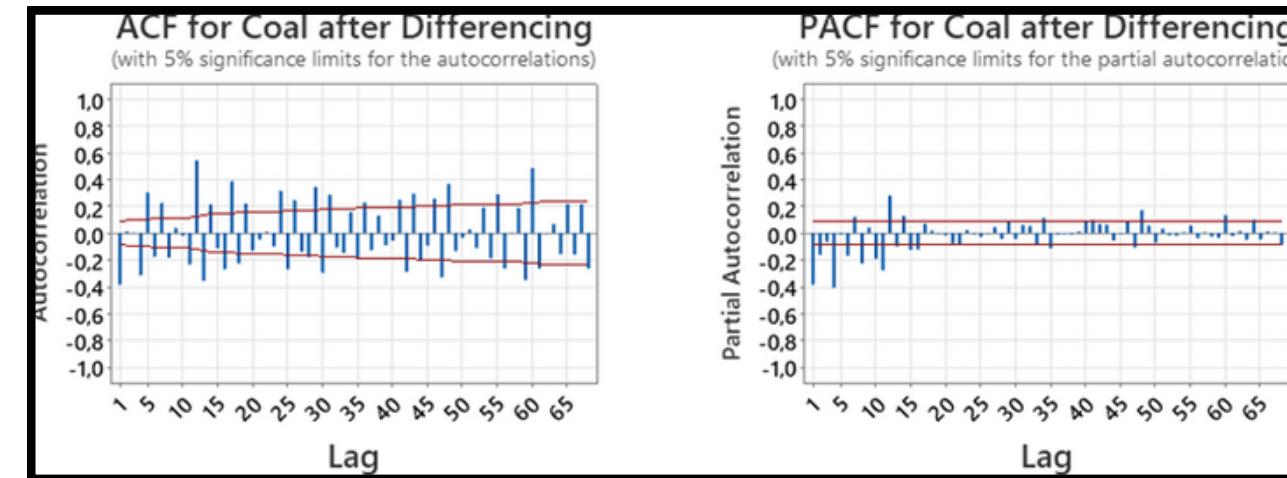
Sebelum Differencing



Menggunakan fitur *Forecast with ARIMA Best Model* yang tersedia dalam perangkat lunak Minitab guna membantu dalam identifikasi dan pemilihan model SARIMA yang paling sesuai.

Model Terpilih	AIC	BIC
ARIMA (0,1,4)(5,1,0) ¹²	10864	10907
ARIMA (0,1,5)(5,1,0) ¹²	10866	10913
ARIMA (4,1,0)(5,1,1) ¹²	10878	10926

Setelah Differencing



Model prediksi yang **optimal** ditentukan berdasarkan nilai **Akaike's Information Criterion (AIC)** dan **Bayesian Information Criterion (BIC)** terendah.

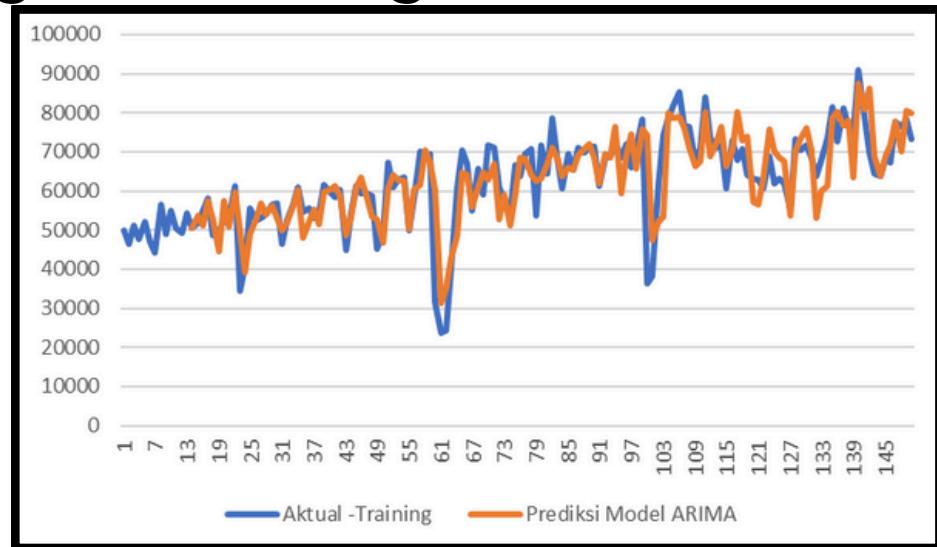
Model Terpilih	RMSE	MAPE
ARIMA (0,1,4)(5,1,0) ¹²	4671,25	4,64%

Hasil dan Pembahasan

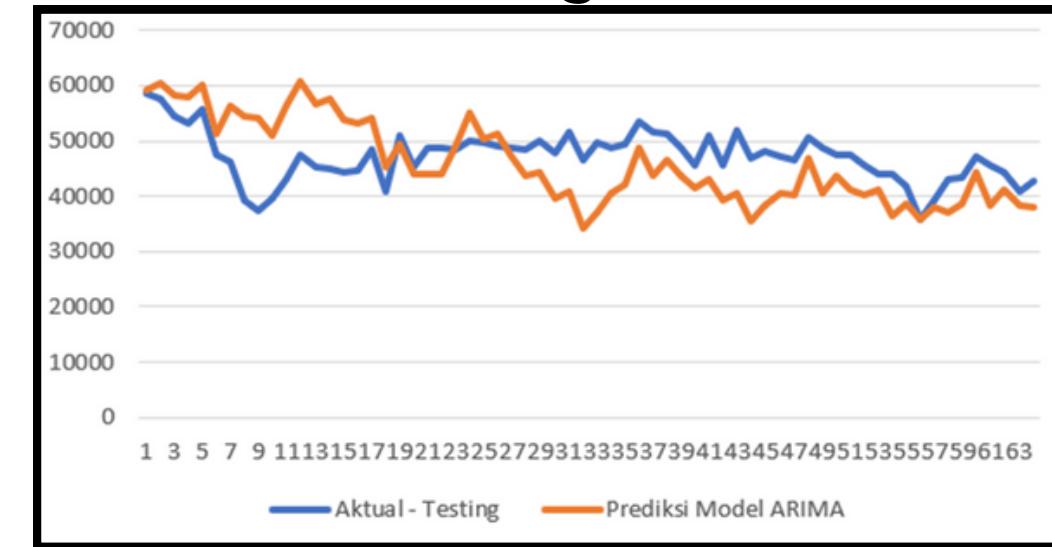
Evaluasi Model SARIMA

Evaluasi model dilakukan melalui dua pendekatan. **Pertama**, dengan melihat **hasil fitting model** terhadap **data pelatihan** untuk menilai seberapa baik model mengenali dan mempelajari pola historis. **Kedua**, dengan menguji kemampuan model dalam **memprediksi data pengujian** yang belum pernah dilihat sebelumnya, guna mengukur tingkat generalisasi.

Plot Fitting Data Training



Plot Prediksi Data Testing



Setelah model dilatih, evaluasi performa dilakukan menggunakan dua metrik utama, yaitu **RMSE** dan **MAPE**, yang digunakan untuk **mengukur tingkat kesalahan prediksi terhadap nilai aktual** dalam bentuk absolut dan persentase.

Dataset	RMSE	MAPE
Training	4671,25	4.64%
Testing	7432,15	13.71%

Hasil dan Pembahasan

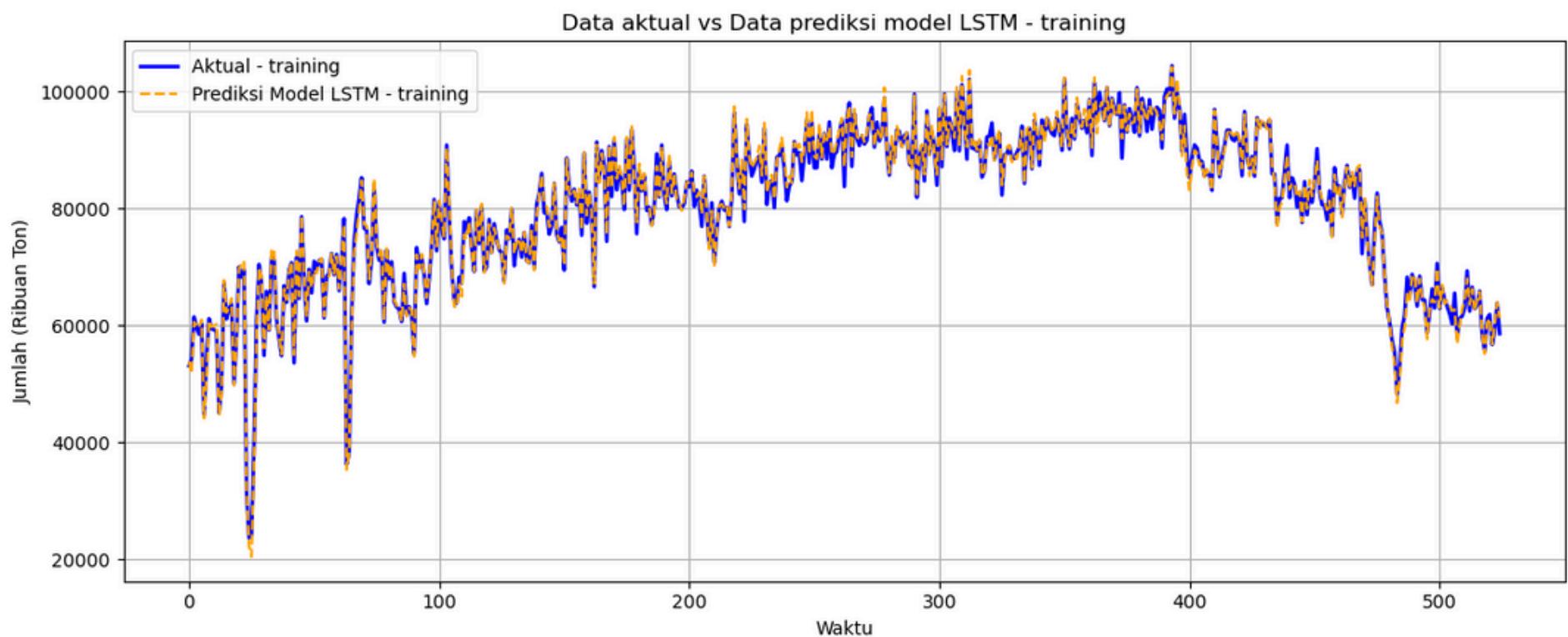
Prediksi dengan Model LSTM

Konfigurasi Model LSTM

Batch Size	32
Sequence Length	36
Hidden Size	256
Number Layers	2
Dropout	0.1
Optimizer	RMSProp
Learning Rate	0.0005
Epochs	300

setelah itu, Model LSTM dilatih menggunakan data historis produksi batu bara dan diuji terhadap data lima tahun terakhir. Hasil prediksi dibandingkan dengan data aktual untuk menilai akurasi model.

Berikut **visualisasi hasil plot** Data Aktual vs Data prediksi model LSTM pada **data training**.

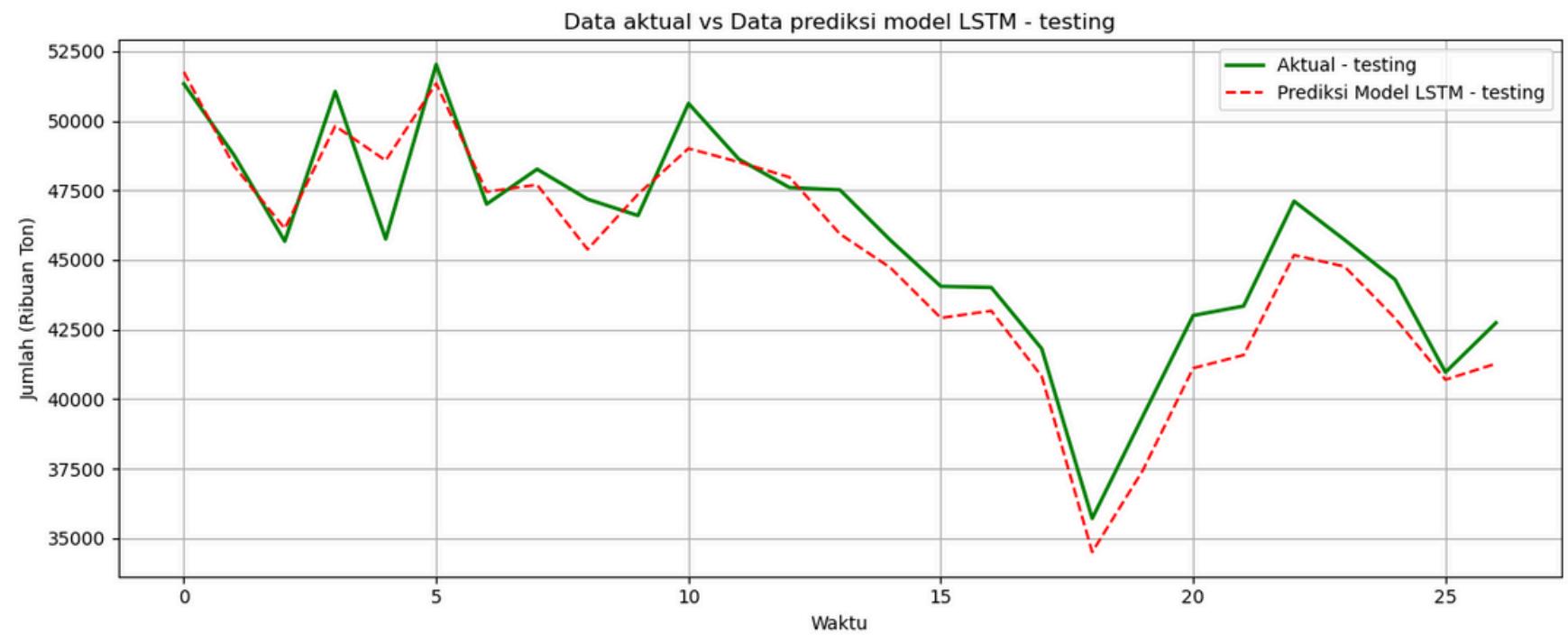


Terlihat bahwa model cukup baik dalam mengikuti pola tren dan fluktuasi musiman, terutama pada periode dengan tren yang relatif stabil.

Hasil dan Pembahasan

Evaluasi Model LSTM

Berikut **visualisasi hasil plot** Data Aktual vs Data prediksi model LSTM pada **data testing**.



Pada data training, model memang mampu mengenali arah tren secara umum, namun masih terdapat beberapa deviasi antara hasil prediksi dan data aktual.

Evaluasi performa dilakukan menggunakan dua metrik utama: **RMSE** dan **MAPE**, untuk mengukur seberapa jauh hasil prediksi menyimpang dari nilai aktual, baik secara absolut maupun persentase.

Dataset	RMSE	MAPE
Training	1113,23	1,14%
Testing	1288,50	2,48%

Nilai **RMSE** dan **MAPE** yang rendah pada data pelatihan menunjukkan **model mampu mempelajari pola historis dengan baik**. Meski terjadi sedikit peningkatan error pada data pengujian, model tetap menunjukkan **generalisasi yang baik terhadap data baru**.

Hasil dan Pembahasan

Evaluasi Model dan Perbandingan

Perbandingan performa prediktif dari kedua model **bertujuan** untuk **menentukan model yang lebih optimal** dalam menangkap pola data dan **menghasilkan prediksi yang mendekati nilai aktual**. Hasil perbandingan nilai kesalahan prediksi antara model SARIMA dan LSTM ditampilkan pada Tabel berikut.

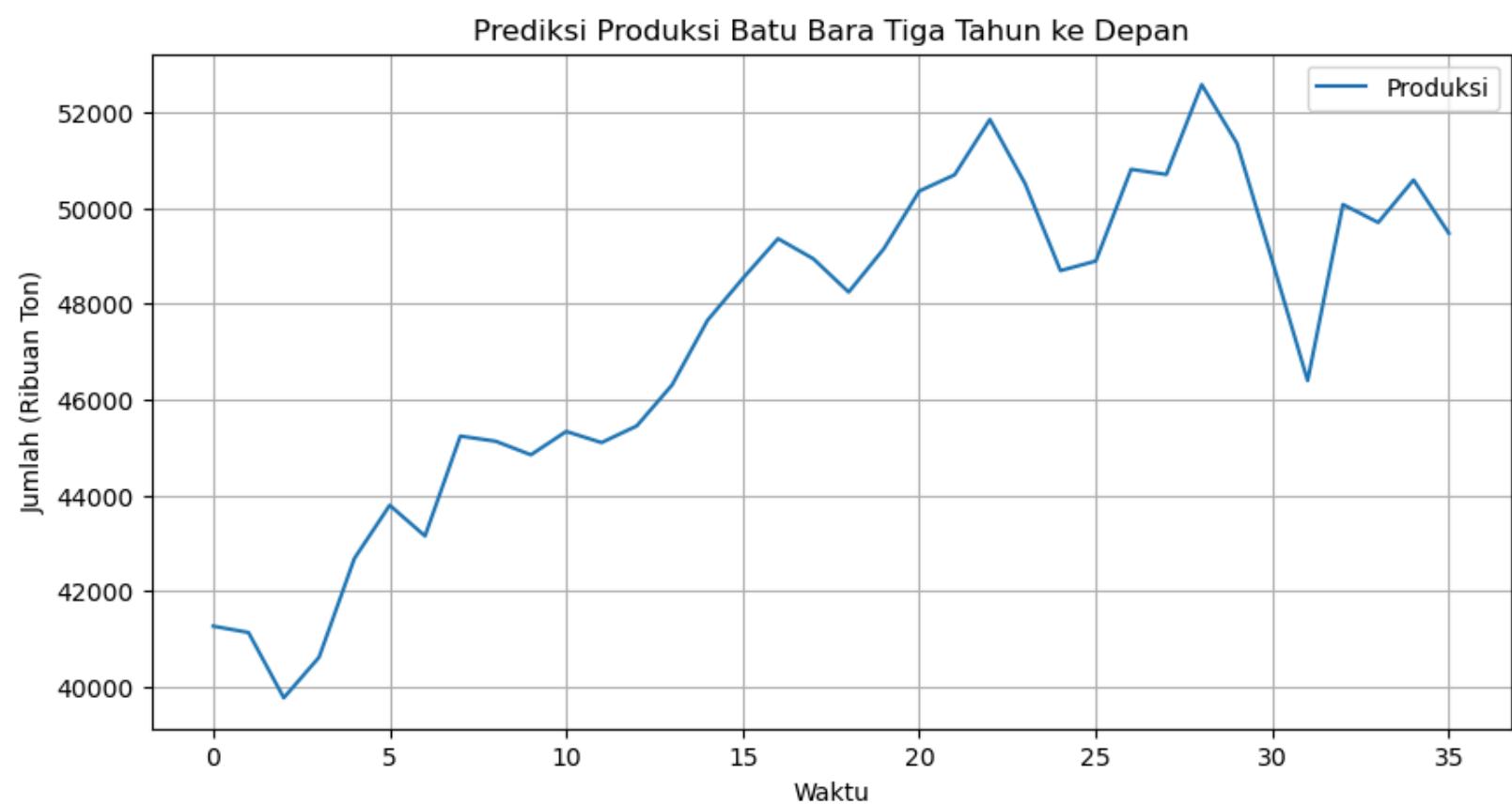
Metode	RMSE	MAPE
ARIMA $(0,1,\underline{4})(5,1,0)^{12}$	7432,15	13,71%
LSTM	1288.50	2.48%

Model **LSTM** secara signifikan menunjukkan **performa yang lebih baik** dibandingkan model **SARIMA**. Nilai **RMSE** dan **MAPE** pada **LSTM jauh lebih rendah**, masing-masing sebesar **1288.50** dan **2.48%**, dibandingkan dengan **SARIMA** yang mencatat **RMSE** sebesar **7432.15** dan **MAPE** sebesar **13.71%**. Hal ini mengindikasikan bahwa **model LSTM** lebih **mampu** menangkap pola **musiman** dan **non linier** dalam data produksi batu bara secara lebih akurat.

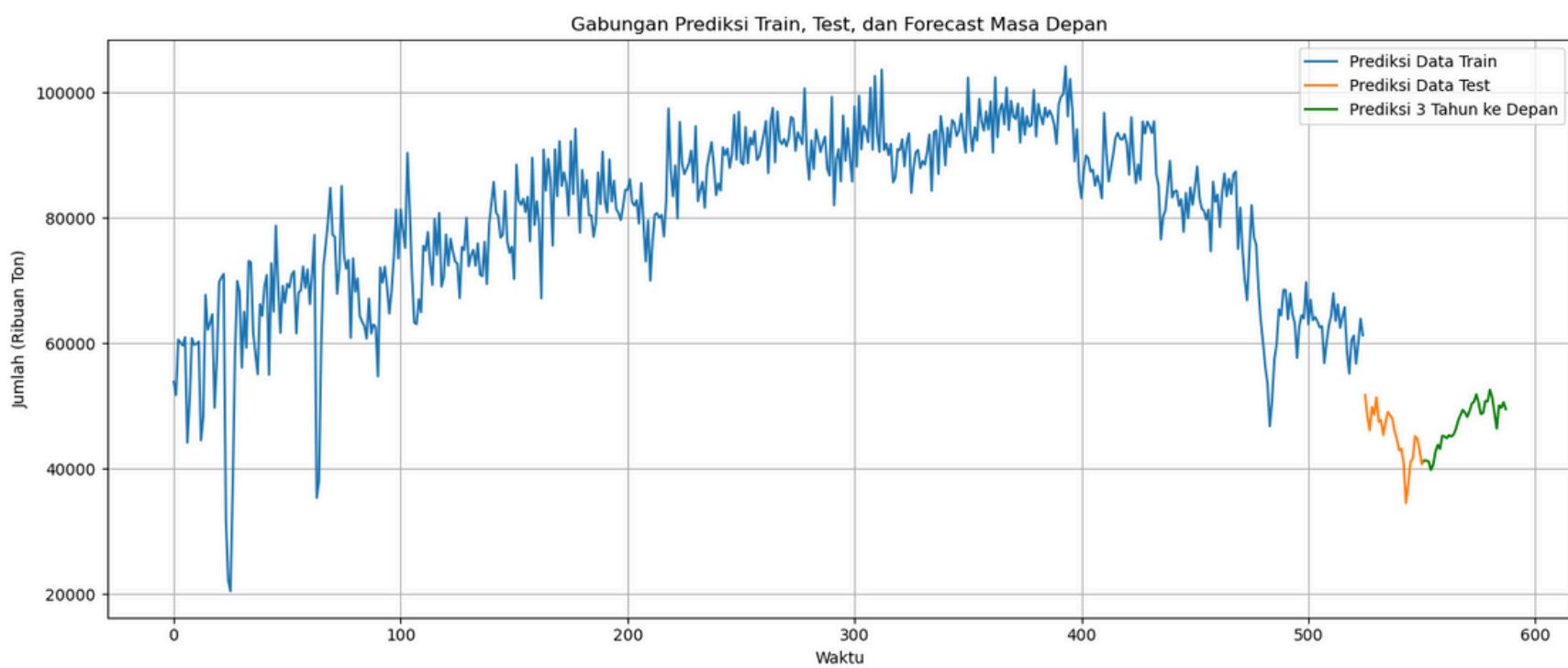
Hasil dan Pembahasan

Prediksi Produksi Tiga Tahun ke Depan

Selanjutnya, Model LSTM digunakan untuk meramalkan produksi batu bara di Amerika Serikat selama tiga tahun ke depan secara autoregresif, dengan input 36 bulan terakhir.



Hasil prediksi menunjukkan pola fluktuatif dengan kecenderungan naik di tahun pertama, lalu naik-turun di tahun kedua dan ketiga, mengikuti pola musiman historis. Berikut plot prediksi pada data pelatihan, pengujian, dan proyeksi tiga tahun ke depan.



KESIMPULAN & SARAN

Kesimpulan

01

Data produksi batu bara di AS memiliki **pola musiman dan tren jangka panjang yang kompleks**.

Uji Terasvirta menunjukkan adanya **nonlinieritas**, sehingga pendekatan seperti **LSTM** lebih sesuai digunakan.

02

Model **ARIMA (0,1,4)(5,1,0)[12]** berhasil menangkap pola **musiman**, namun **kurang fleksibel terhadap pola nonlinier**. Hasil evaluasi menunjukkan **RMSE sebesar 7.432,15** dan **MAPE 13,71%** pada data pengujian.

03

Prediksi **LSTM dua tahun** ke depan menunjukkan **pola stabil dan musiman**, dengan produksi diperkirakan 48.000–52.000 ribu ton per bulan. Hasil ini dapat jadi dasar kebijakan energi.

04

Model **LSTM** menunjukkan hasil prediksi yang **jauh lebih akurat**. Setelah dilatih, model ini menghasilkan **kesalahan prediksi yang rendah**. Hal ini menunjukkan bahwa **LSTM mampu memahami pola data yang kompleks** dan memberikan hasil yang lebih andal untuk peramalan produksi batu bara.

05

Berdasarkan perbandingan, **LSTM lebih unggul dari SARIMA** dalam meramalkan data produksi batu bara yang kompleks dan tidak linier. **LSTM mampu menangkap pola yang rumit**, sementara **SARIMA** lebih cocok untuk data yang **stabil dan linier**. Namun, **LSTM** membutuhkan **komputasi lebih besar** dan penyetelan model yang lebih rumit, sedangkan **SARIMA** lebih **sederhana** dan **mudah dipahami**.



Saran

Penelitian Selanjutnya

- **Pemodelan LSTM** dapat diperluas dengan **menambahkan fitur-fitur eksternal (multivariat)**, seperti harga energi, konsumsi, atau cuaca, untuk meningkatkan akurasi prediksi dan mengakomodasi pengaruh faktor eksternal.
- Eksperimen lanjutan dengan **model hybrid** seperti **kombinasi SARIMA-LSTM** atau penggunaan **model Transformer Time Series** dapat dieksplorasi sebagai pendekatan alternatif untuk menangkap pola linier dan non-linier secara bersamaan
- Penggunaan **optimasi hiperparameter otomatis** seperti **PSO** dapat meningkatkan efektivitas model LSTM secara signifikan, sehingga tidak hanya mengandalkan konfigurasi manual.

Pemerintah AS

- **Hasil penelitian** ini diharapkan dapat menjadi **rujukan bagi pemerintah Amerika Serikat**, pelaku industri, dan perencana energi dalam **menyusun strategi berbasis data yang adaptif terhadap perubahan dinamika sektor energi**, khususnya dalam pengelolaan produksi batu bara nasional.
- **Prediksi jangka menengah** yang dihasilkan dari model ini dapat **dimanfaatkan** oleh instansi terkait seperti U.S. Energy Information Administration dan Department of Energy sebagai **bahan perencanaan produksi dan pengawasan rantai pasok energi nasional**. Namun demikian, pembaruan model secara berkala tetap diperlukan untuk menjaga akurasi seiring bertambahnya data baru



Thank You!
for the attention