

FORECASTING

● GRUP 2

Aris Chandra
Anita Dilli
Anwar Fauzi

TABLE OF CONTENT

CASE : PENJUALAN BUKU DI BULAN DEPAN

01

Analisis Tren Bulanan penjualan buku berdasarkan data historis

02

Gunakan metode peramalan yang lebih cocok untuk peramalan jangka panjang, seperti metode time series (misalnya, SARIMA) atau metode ensemble (misalnya, Random Forest).

03

Lakukan prediksi penjualan buku di bulan depan dan berikan rekomendasi stok ke tim produksi dan tim gudang.



DAFTAR BUKU TERLARIS

BELAJAR MS OFFICE DENGAN MUDAH DAN MENYENANGKAN

BUDIDAYA IKAN SISTEM BIOFLOK

PENERAPAN MODEL PEMBELAJARAN
BERDIFERENSIASI

MODEL-MODEL PEMBELAJARAN

DASAR-DASAR
TEKNIK INFORMATIKA

ILMU LOGIKA

TEKNIK PENULISAN
KARYA ILMIAH

0

500

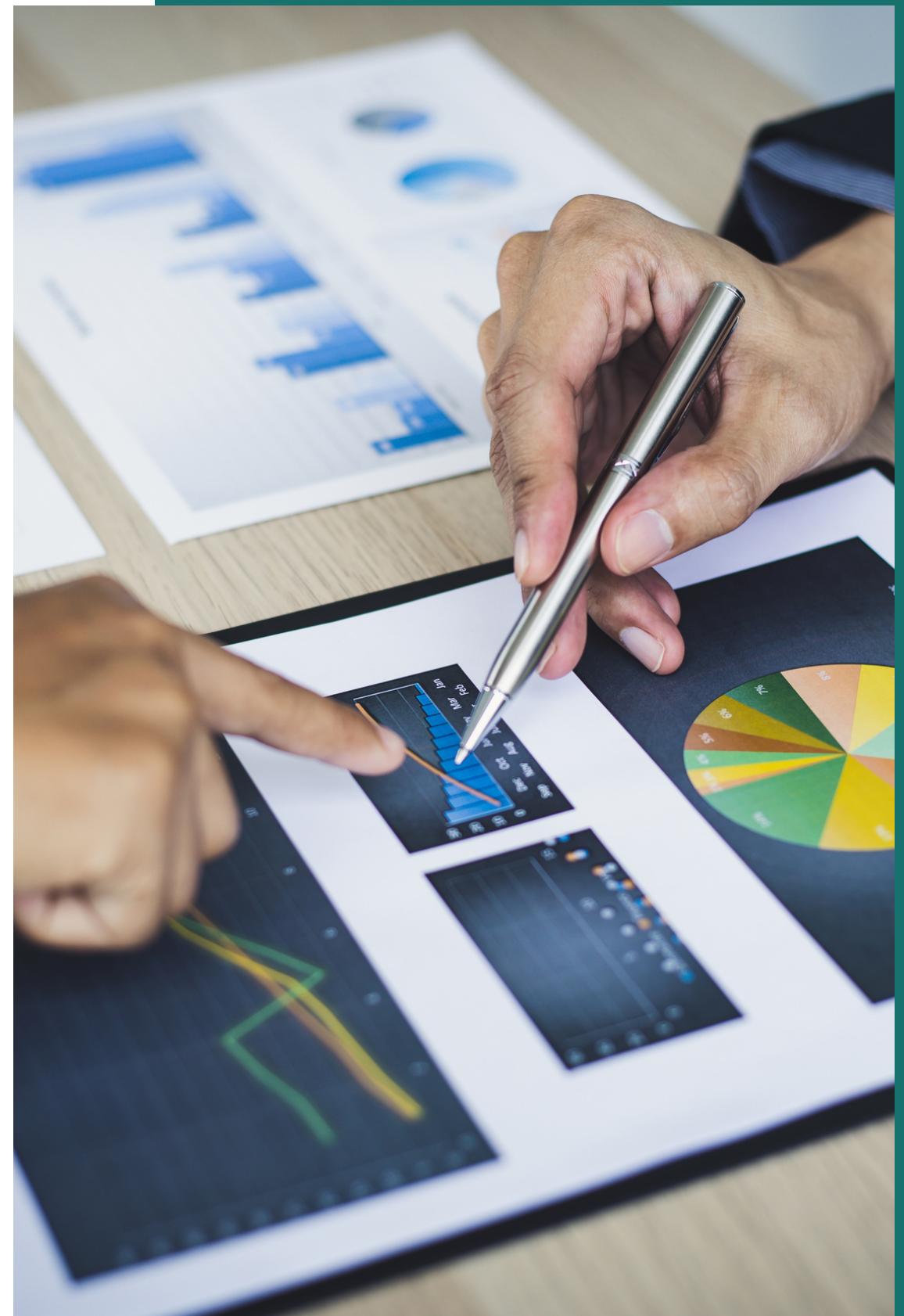
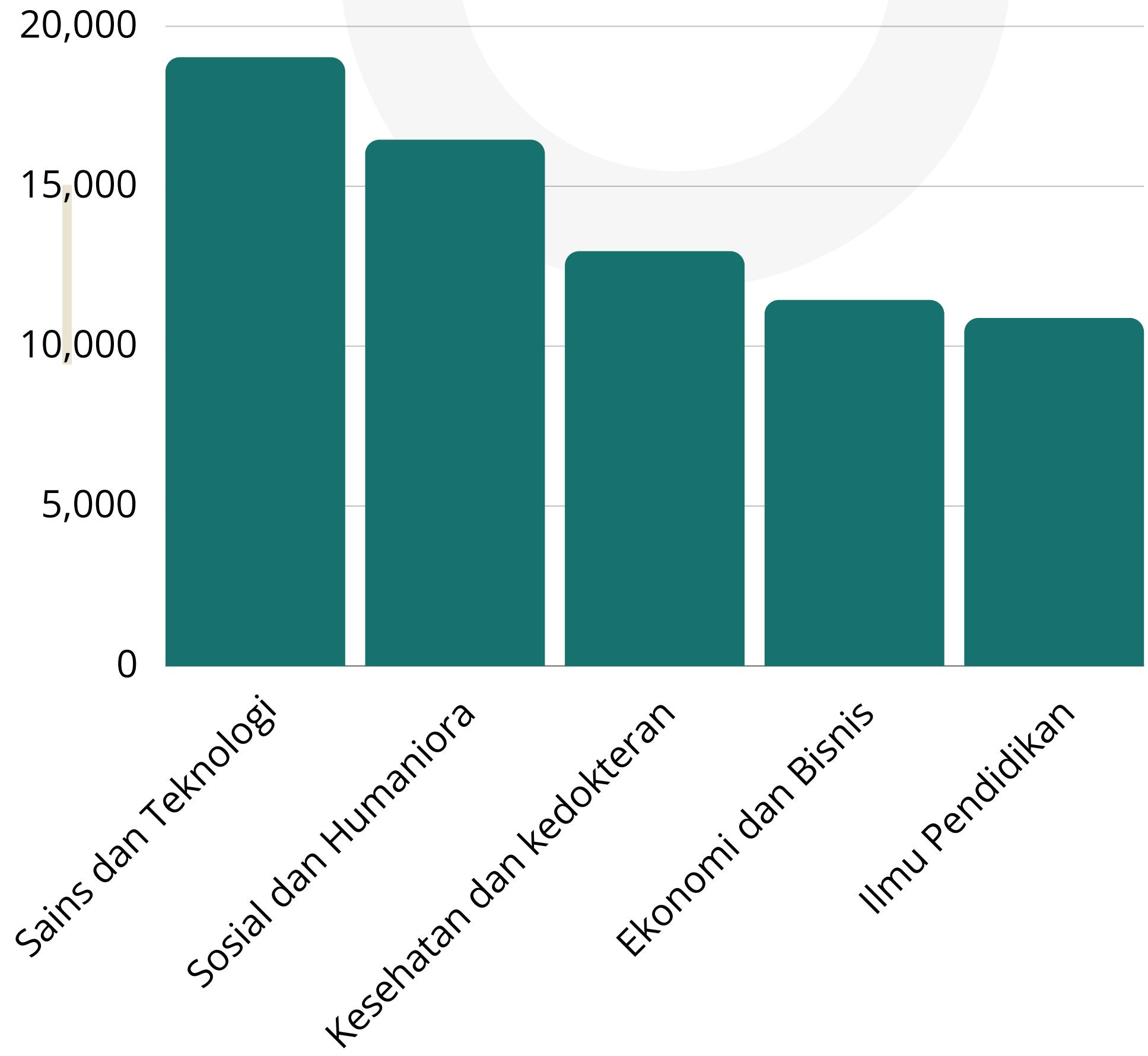
1,000

1,500

2,000

BERDASARKAN ANALISIS KAMI, BUKU BELAJAR MICROSOFT OFFICE (WORD, EXCEL, POWERPOINT) 2019 DENGAN MUDAH DAN MENYENANGKAN DENGAN TOTAL PENJUALAN 1976 EKSEMPLAR.

DISTRIBUSI PENJUALAN BUKU BERDASARKAN KATEGORI



DISTRIBUSI PENULIS BUKU TERKAIT PENJUALAN



Johnie Roger Swanda Pasribu

Ajat Rukajat

Shilphy A Octavia

Banyumi

Gusrina

Kurniawan

0

200

400

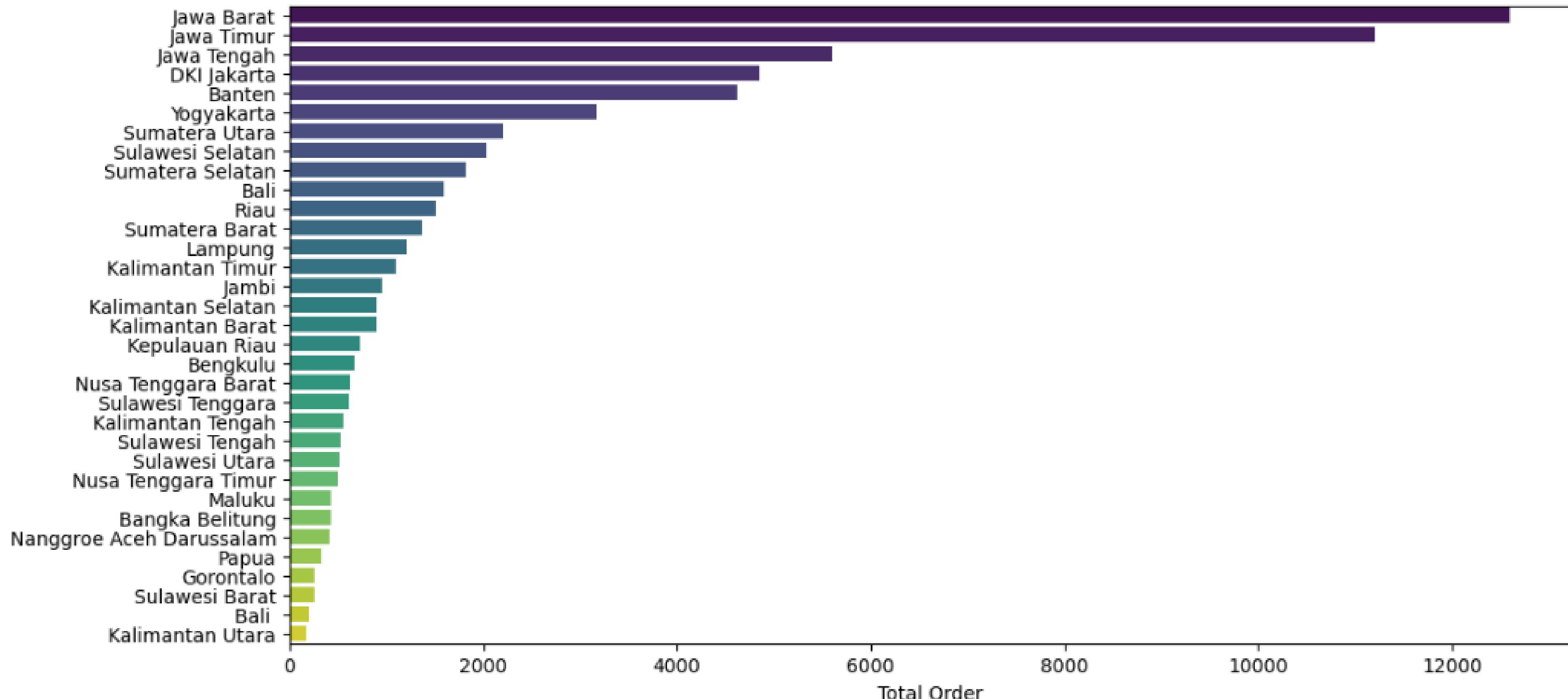
600

800

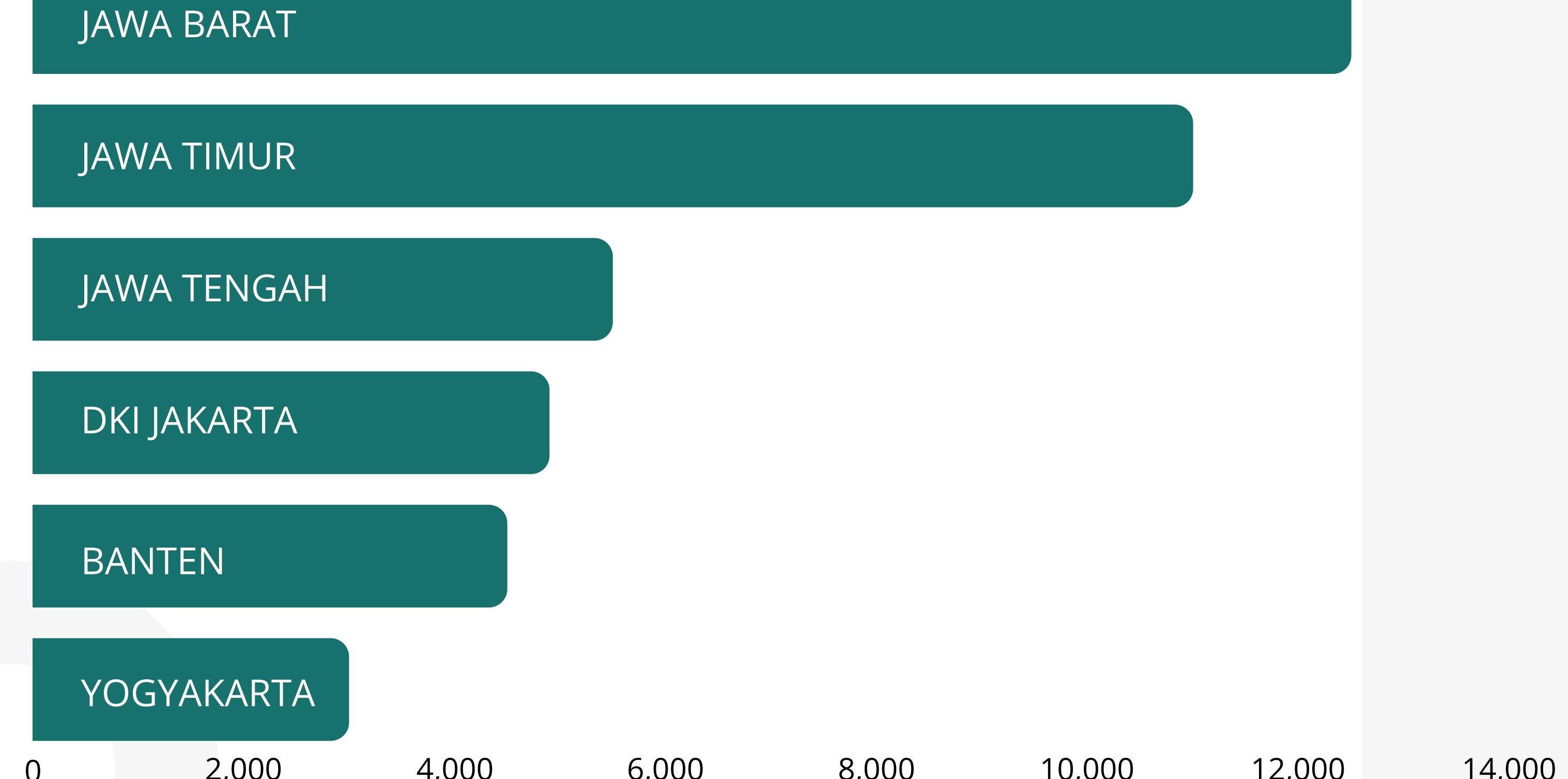
1,000

DISTRIBUSI ORDER BERDASARKAN PROVINSI

Distribusi Order berdasarkan Provinsi



DISTRIBUSI ORDER BERDASARKAN PROVINSI



01

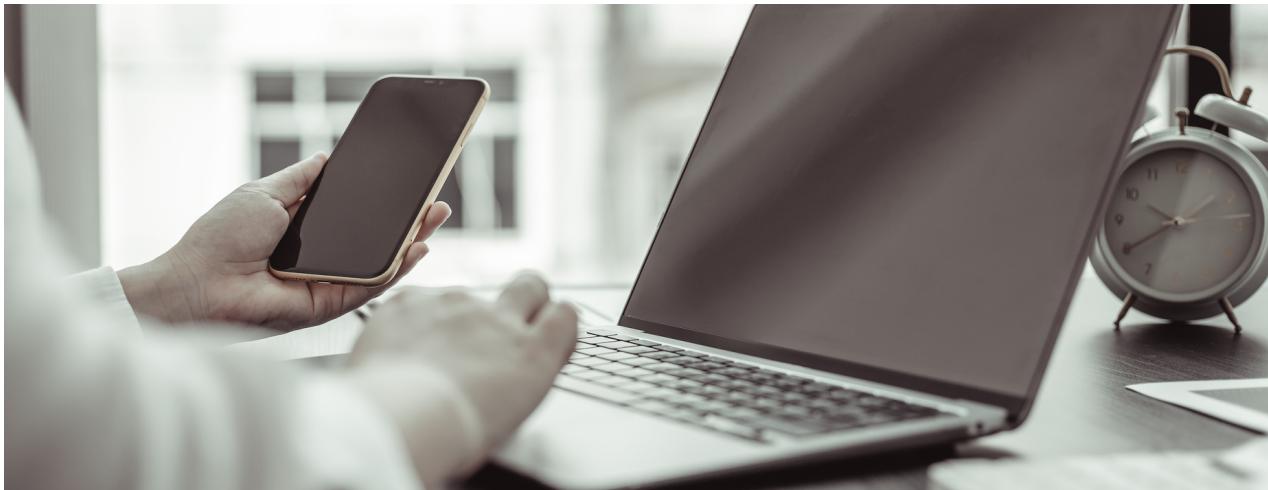
5 Provinsi dengan Penjualan buku tertinggi berada di Pulau jawa, karena ada promo diskon ongkir setiap pembelian buku dengan minimal transaksi diatas Rp80.000, sehingga banyak pengiriman buku di pulau jawa tinggi

02

Berdasarkan hasil analisa kami, ada korelasi antara penjualan Kategori penjualan buku terbanyak yaitu Kategori Sains dan Teknologi dengan penulis buku terlaris yaitu Johnie Rogers. Sehingga adanya pengaruh penjualan buku kategori sains terhadap penulis buku terlaris.

02

Deepublish bisa melakukan promosi dengan maksimal di pulau jawa, karena mayoritas penjualan tertinggi ada di pulau jawa yang merupakan setengah penduduk indonesia ada di pulau jawa.



BUKU TERLARIS DI JAWA BARAT

Belajar MS Office dengan mudah dan menyenangkan

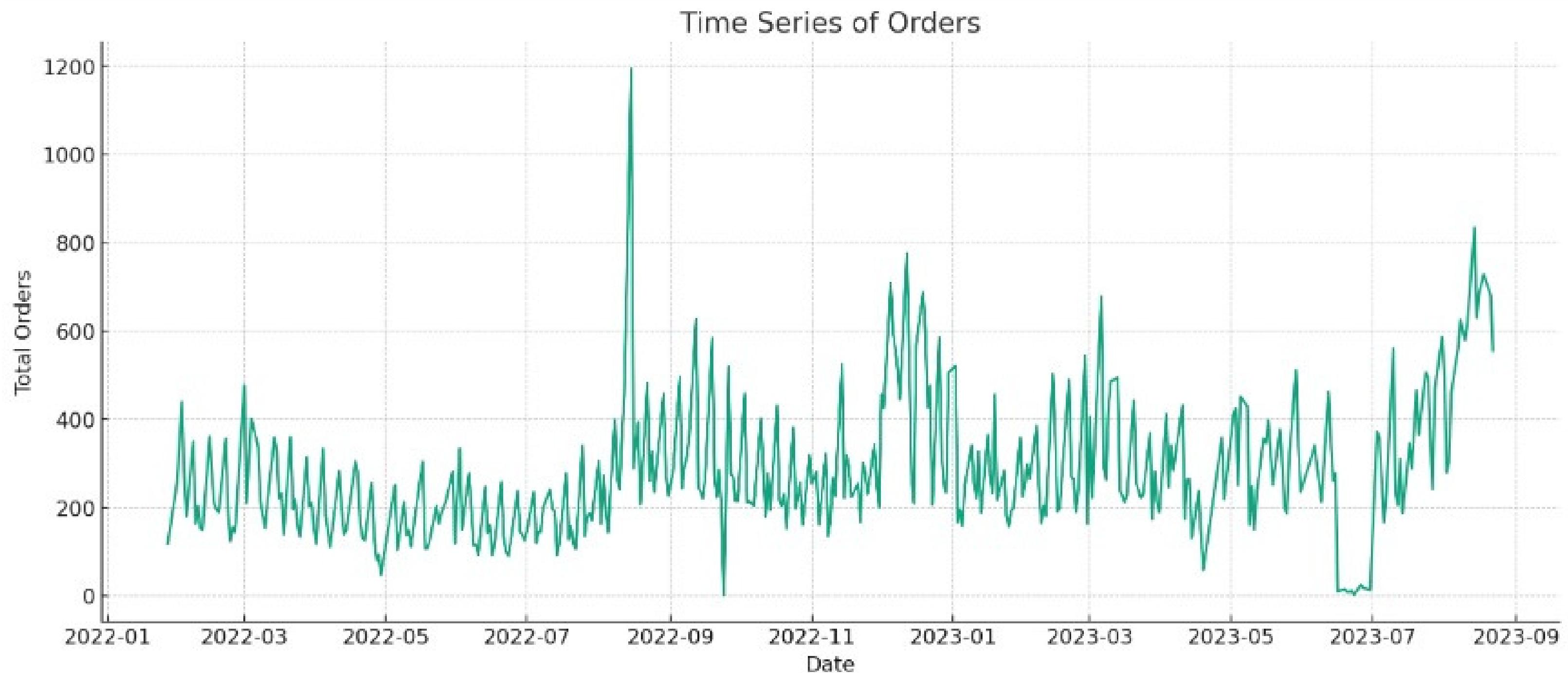
Teknik Penulisan Karya Ilmiah

Pendidikan Keachmadiyanian

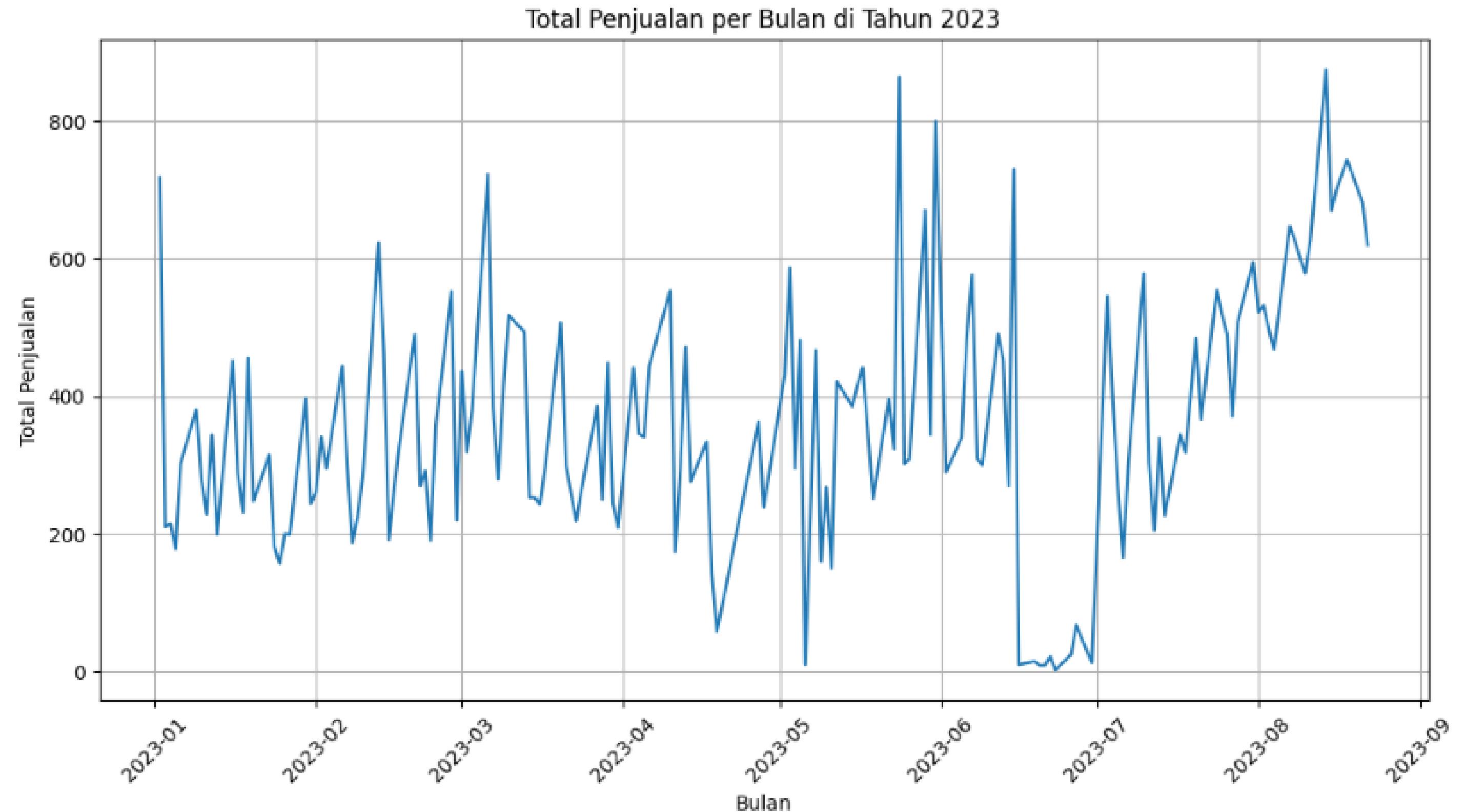
Budidaya ikan sistem Bioflok

Sekolah sambil kerja
di Australia

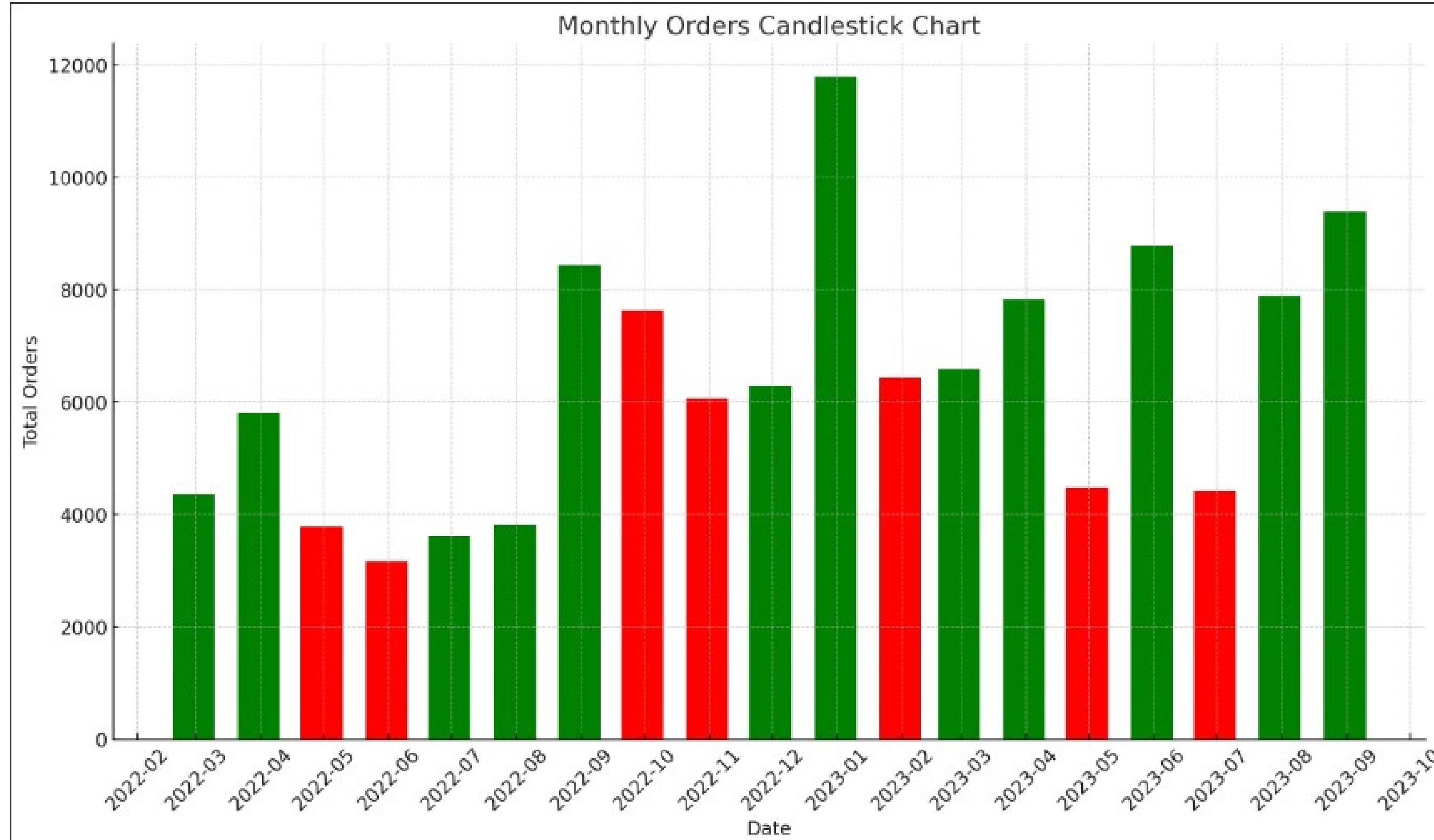
ANALISIS TIME SERIES



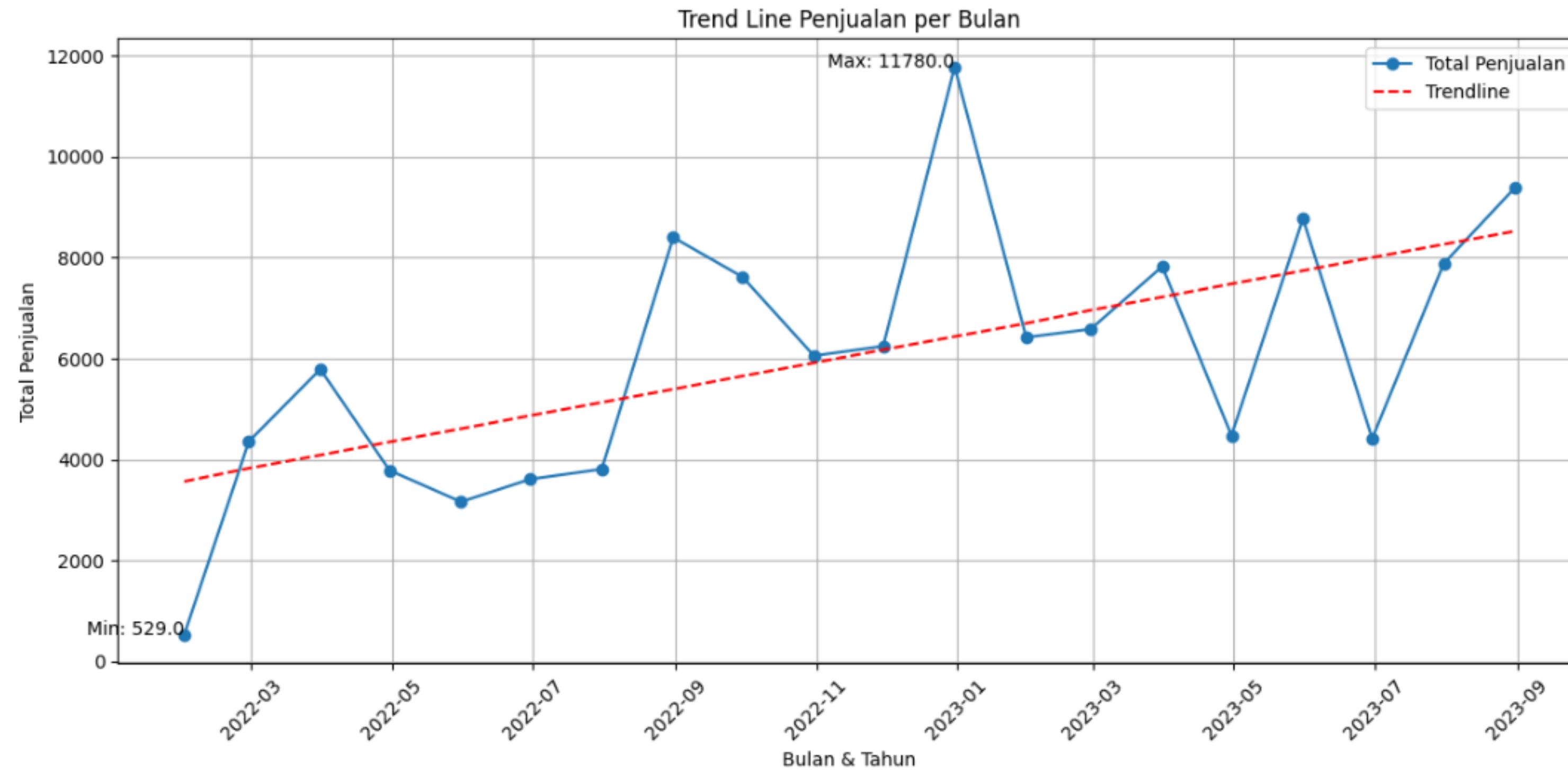
ANALISIS TIME SERIES



TIME SERIES RESULT



UPTREND ANALYSIS



STATIC METHODS DISKEY-FULLER

```
[ ] def adf_test(timeseries):
    print ('Results of Dickey-Fuller:')
    print ('-----')
    adftest = adfuller(timeseries)
    adf_output = pd.Series(adftest[0:4], index=['Test Statistic','p-value','#Lags Used','Number of Observations Used'])
    for key,value in adftest[4].items():
        adf_output['Critical Value (%s)'%key] = value
    print (adf_output)

# calling adf function and passing series
adf_test(dfnew_order.values)
```

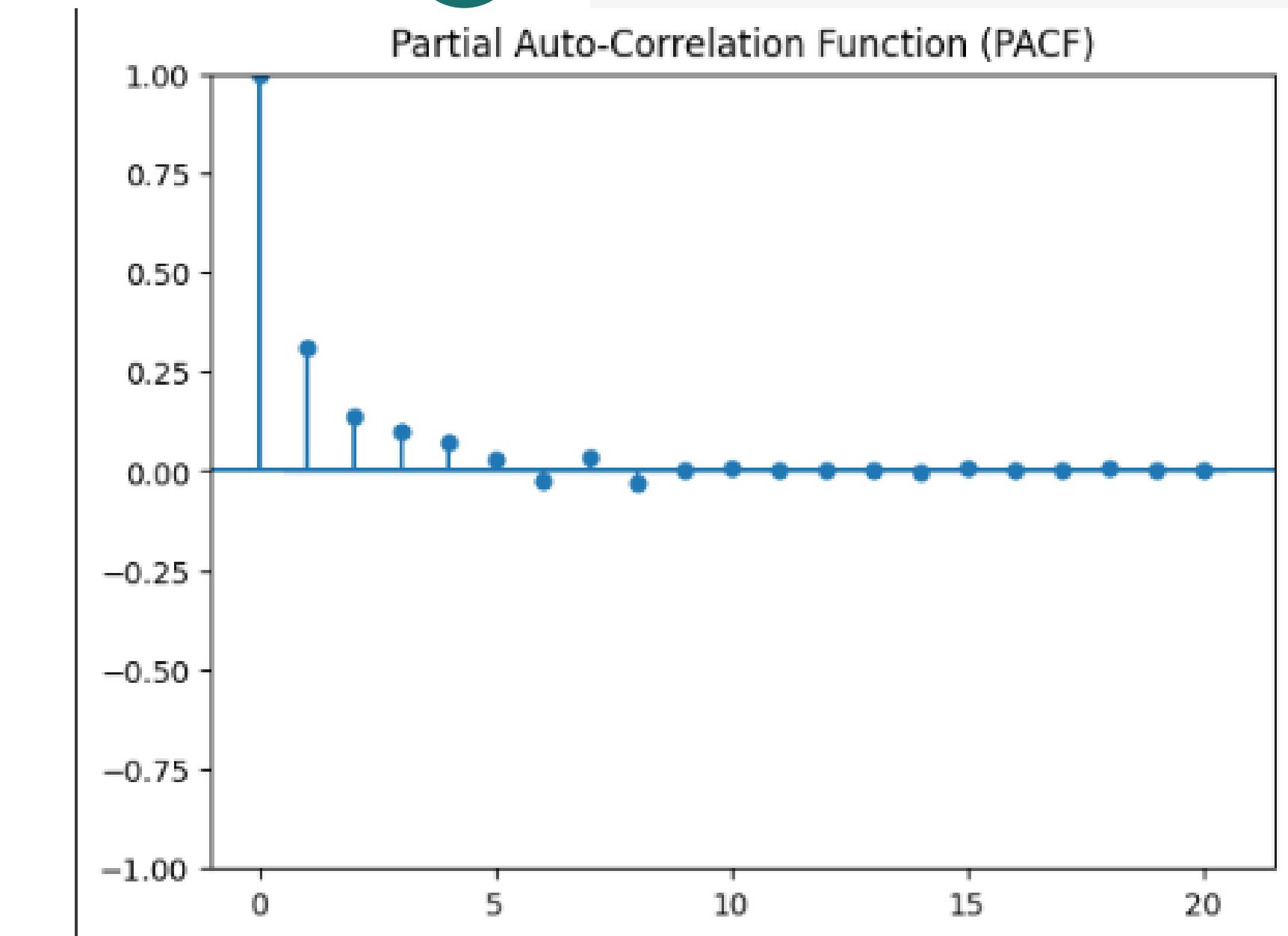
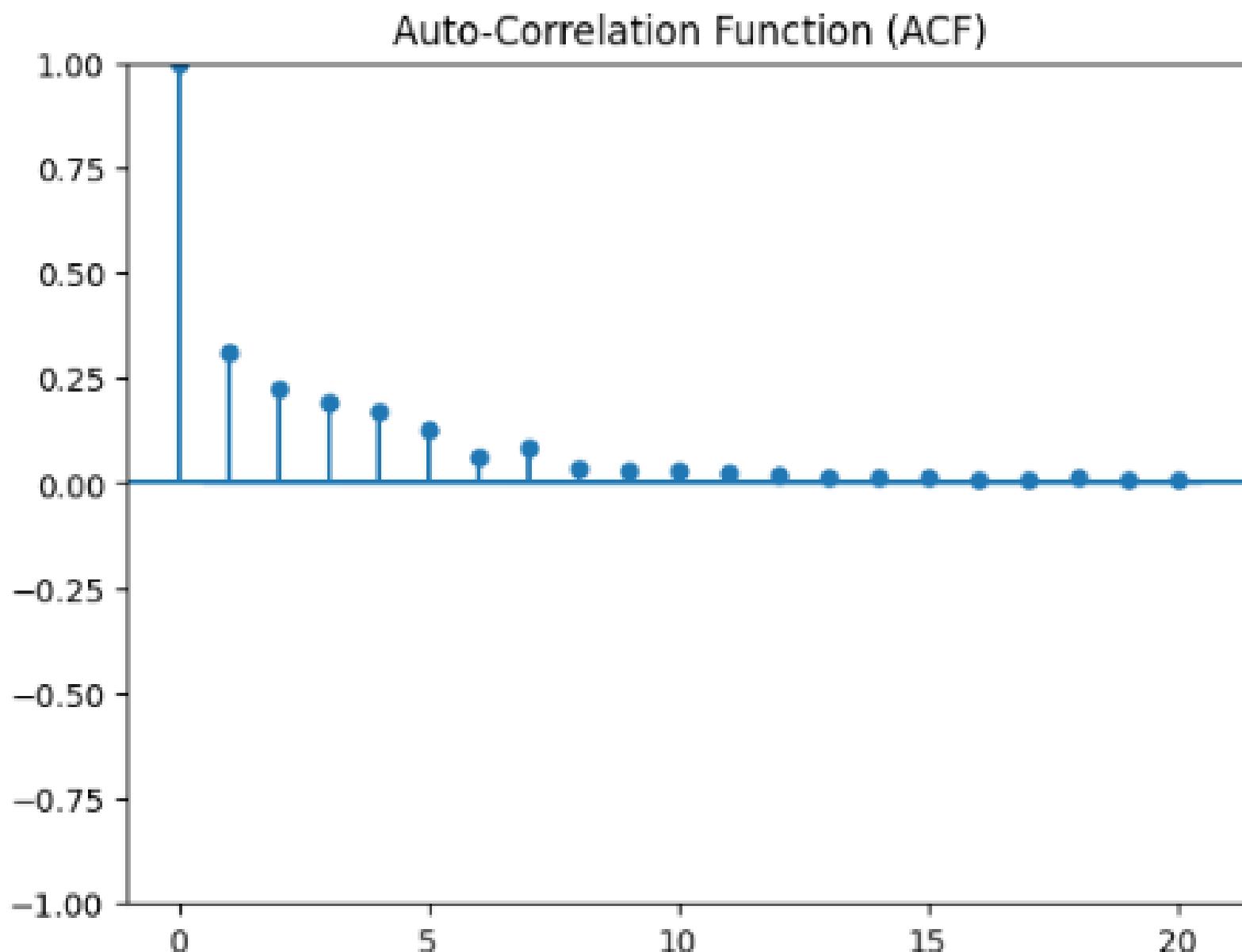
Results of Dickey-Fuller:

```
-----  
Test Statistic           -77.738588  
p-value                  0.000000  
#Lags Used              9.000000  
Number of Observations Used 87003.000000  
Critical Value (1%)      -3.430425  
Critical Value (5%)      -2.861573  
Critical Value (10%)     -2.566788  
dtype: float64
```

Hasil uji Dickey-Fuller:

Berdasarkan hasil uji Dickey-Fuller, kita dapat melihat bahwa p-value = 0.000000. P-value ini jauh lebih kecil dari nilai kritis Alpha (0.05), Ini berarti bahwa hasil uji Dickey-Fuller tidak mungkin terjadi secara kebetulan jika data tidak stasioner sehingga kita dapat menyimpulkan bahwa data Deepublish adalah stasioner. Data stasioner adalah data yang tidak memiliki tren atau musiman. Data yang stasioner diperlukan untuk membuat model SARIMA. Oleh karena itu, hasil uji Dickey-Fuller ini menunjukkan bahwa data ini dapat digunakan untuk membuat model SARIMA.

IDENTIFIKASI KORELASI



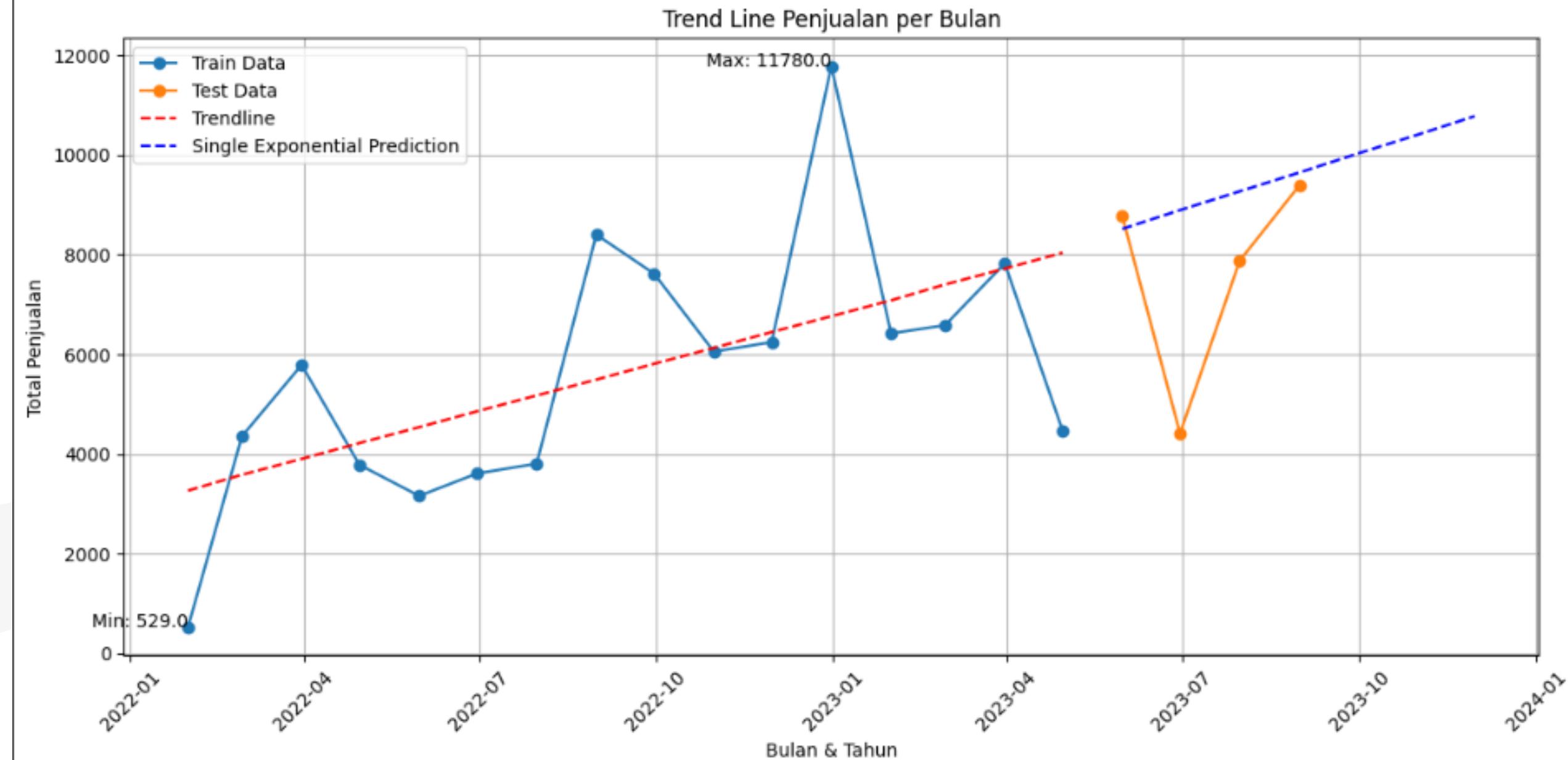
Berdasarkan grafik ACF dan PACF , data deret waktu ini memiliki pola musiman dengan periode 2. Oleh karena itu, model SARIMA yang tepat untuk data deret waktu ini adalah model SARIMA(2,0,0) (2,0,0)12.



Single Exponential Smoothing



```
Train RMSE: 2059.4492048355537  
Test RMSE: 2351.4076940254395  
Train MAPE: 27.605376416099723  
Test MAPE: 20.94726106365861
```

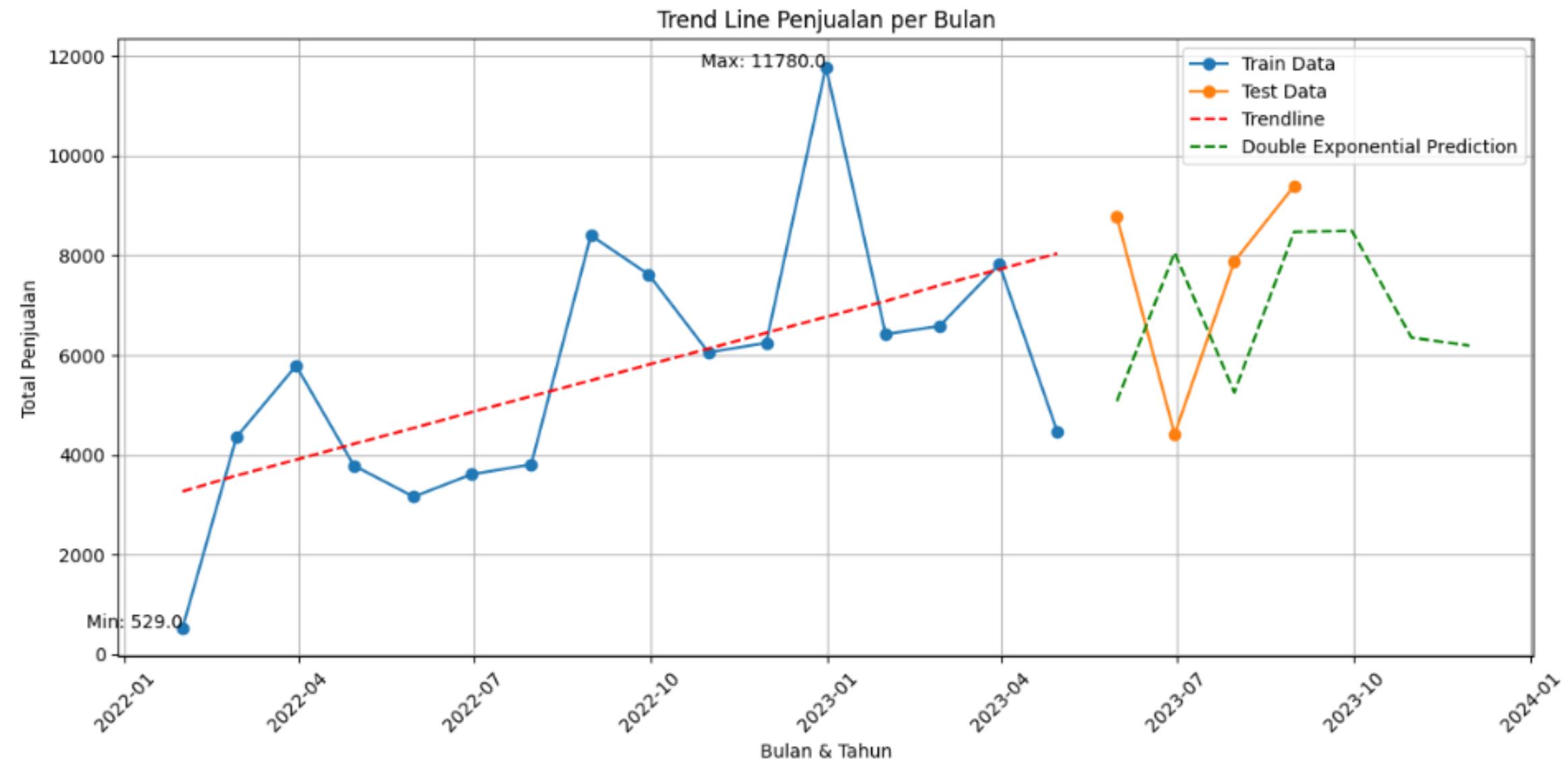


RMSE (Root Mean Squared Error) dan MAPE (Mean Absolute Percentage Error) adalah matrik evaluasi kinerja model regresi, khususnya dalam konteks ini, untuk memeriksa seberapa baik model regresi dan prediksinya cocok dengan data aktual.

Doubel Exponential Smoothing

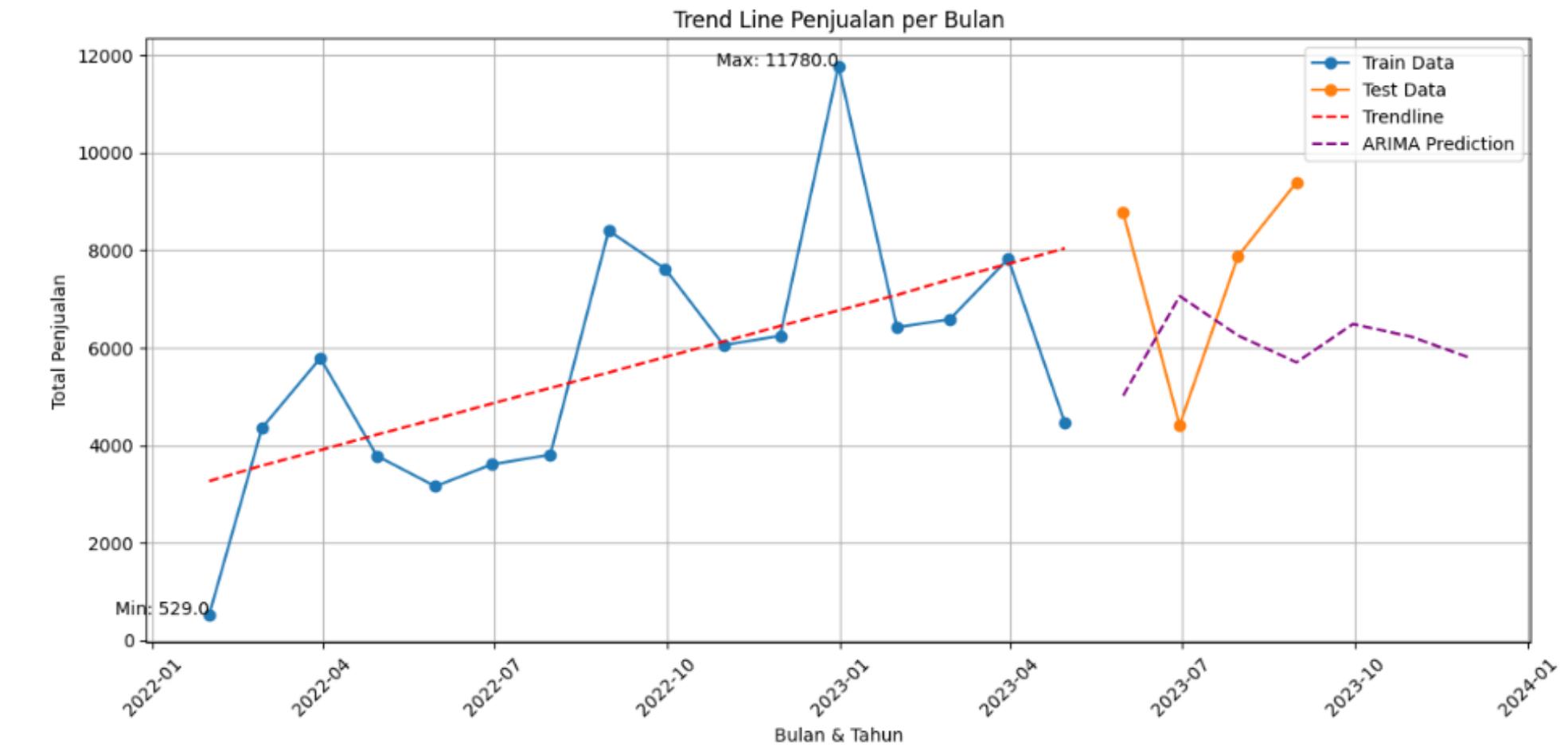
Nilai Root Mean Squared Error (RMSE) untuk data pelatihan menggunakan model double eksponensial. RMSE mengukur seberapa baik model cocok dengan data aktual, dan nilai yang lebih rendah menunjukkan kesesuaian yang lebih baik.

Nilai MAPE yang lebih tinggi (35.71%) pada data pengujian menunjukkan bahwa model mungkin kurang akurat dalam memprediksi data yang tidak digunakan selama pelatihan. Secara keseluruhan, hasil RMSE dan MAPE tersebut memberikan gambaran tentang kinerja model double eksponensial. Nilai-nilai tersebut dapat membantu dalam mengevaluasi kecocokan model terhadap data pelatihan dan pengujian serta memberikan pemahaman seberapa baik model dapat digunakan untuk prediksi.



ARIMA

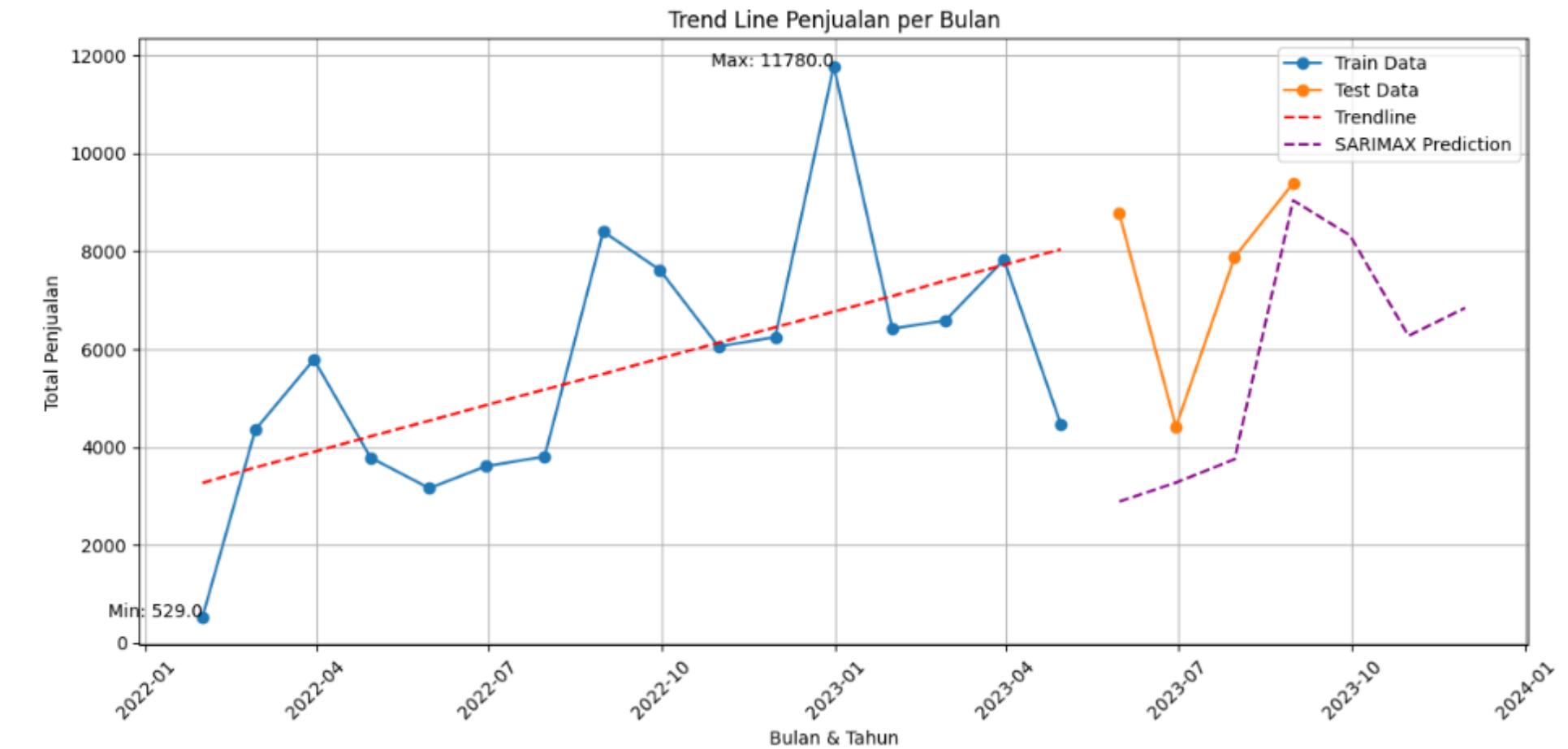
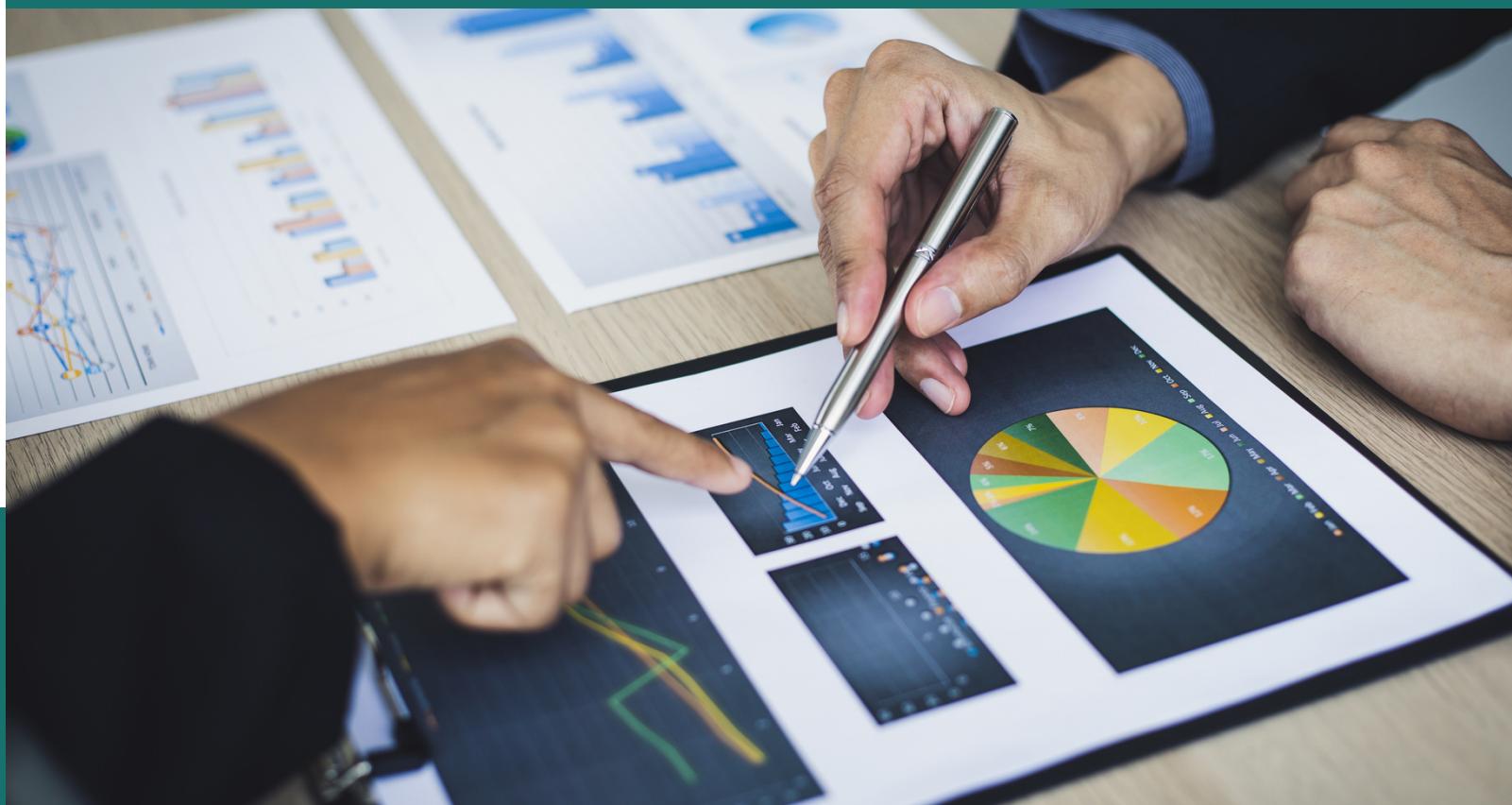
Model ARIMA mungkin tidak sepenuhnya memodelkan tren dan pola dalam data, terutama ketika dihadapkan pada data yang belum pernah dilihat sebelumnya. Evaluasi model ini memberikan wawasan tentang sejauh mana model dapat diandalkan dalam memprediksi nilai masa depan dan menyoroti area di mana perbaikan atau model alternatif mungkin perlu dipertimbangkan.



```
ARIMA Train RMSE: 2415.954447244186
ARIMA Train MAE: 1690.7952631183994
ARIMA Train MAPE: 29.903416947499515
ARIMA Test RMSE: 3056.889193937631
ARIMA Test MAE: 2930.2235772633285
ARIMA Test MAPE: 38.472048542812686
```

SARIMA

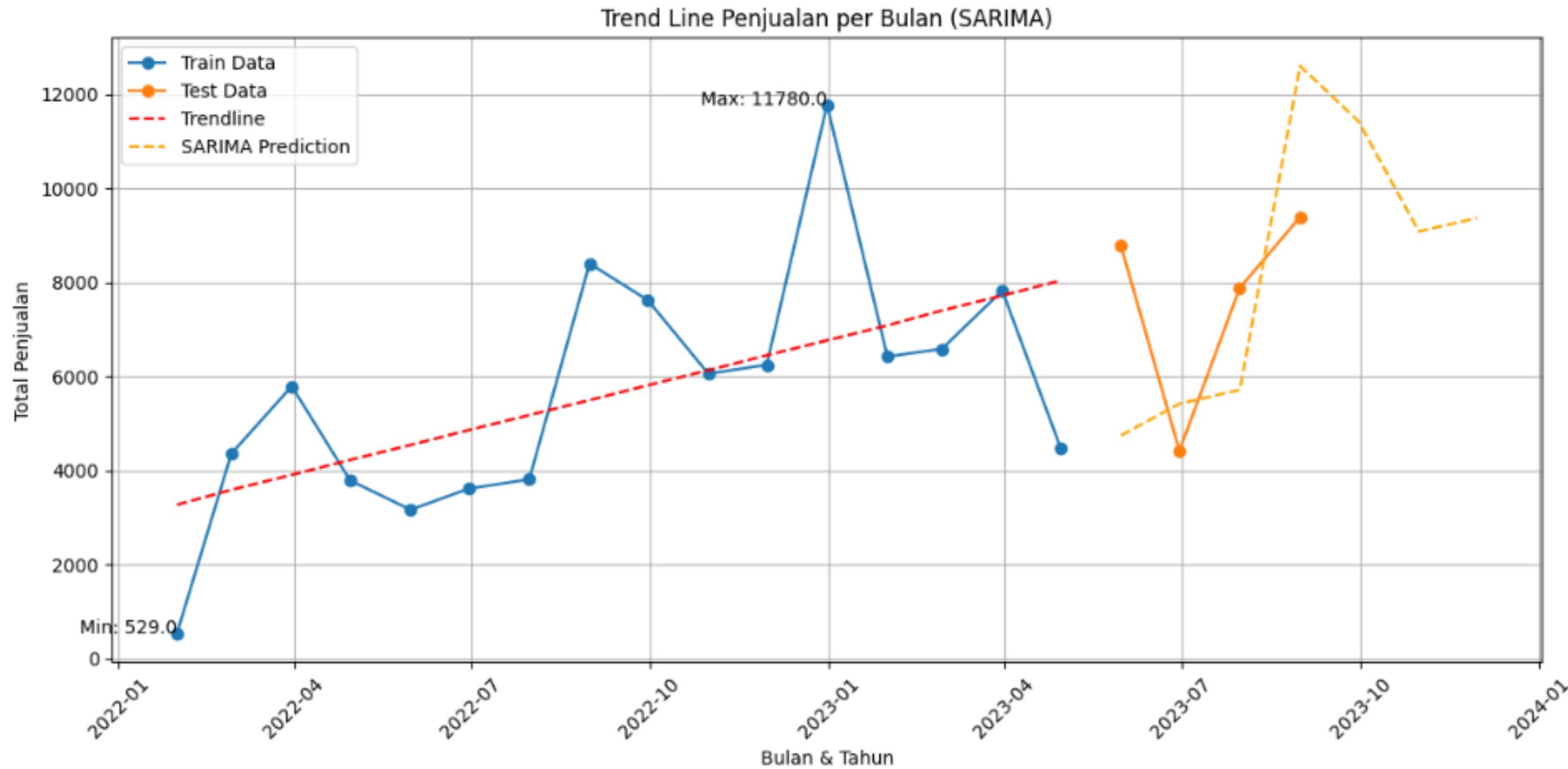
Berdasarkan hasil evaluasi di atas, model SARIMA tampaknya memberikan performa yang baik dalam memprediksi penjualan per bulan. Meskipun terdapat kesalahan dalam prediksi, nilai RMSE dan MAPE yang relatif rendah menunjukkan bahwa model ini dapat diandalkan untuk digunakan dalam meramalkan tren penjualan di masa depan.



SARIMAX Train RMSE: 3131.5188661999605
SARIMAX Test RMSE: 3643.036208807938
SARIMAX Train MAPE: 34.56283756168085
SARIMAX Test MAPE: 37.76076091242444

TUNING MODEL SARIMA

Berdasarkan hasil Tuning terhadap model Sarima, Sangat mengejutkan melihat bahwa nilai-nilai terbaik yang dihasilkan adalah $(0, 0, 0)$ untuk orde non-musiman dan $(0, 2, 0, 12)$ untuk orde musiman. Biasanya, model SARIMA yang lebih kompleks memiliki kinerja yang lebih baik. Namun, kita akan menggunakan parameter terbaik yang ditemukan untuk melihat visualisasi prediksi



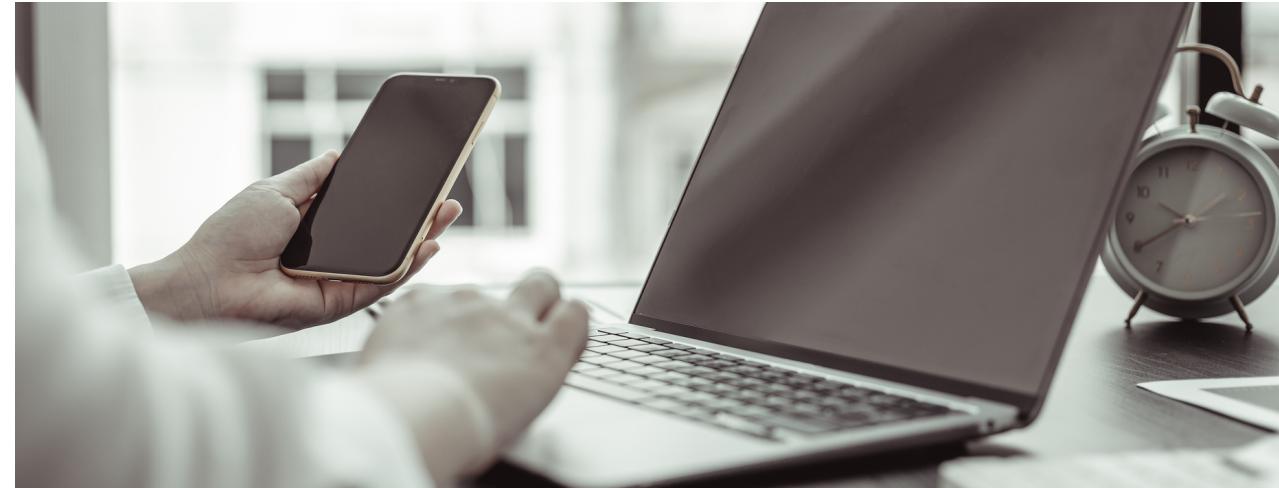
```
warn('Too few observations to estimate starting parameters.')
Best Order: (0, 0, 0)
Best Seasonal Order: (0, 2, 0, 12)
Best AIC: 2.0
```

Hasil Tuning Model Sarimax

Best Order: (0, 0, 0) Best Seasonal Order: (0, 2, 0, 12) Best AIC: 2.0

SARIMAX Train RMSE: 3131.5188661999605
SARIMAX Test RMSE: 3643.036208807938
SARIMAX Train MAPE: 34.56283756168085
SARIMAX Test MAPE: 37.76076091242444

METODE TIME SERIES SARIMA



01

Terjadi Trend Peningkatan

02

Trend Peningkatan berulang / Historikal

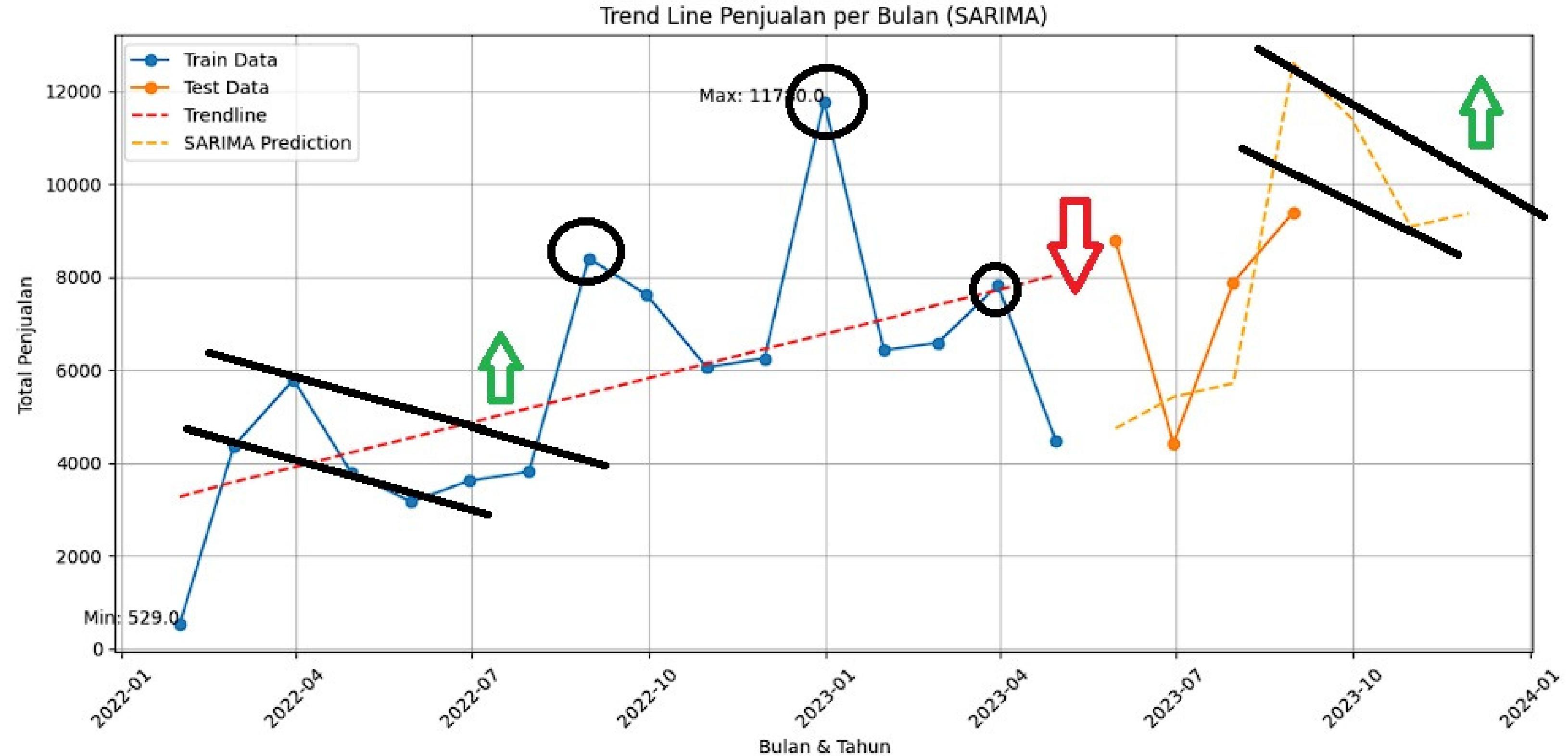
Contoh: pada bulan 8/22 ke 9/22 terjadi kenaikan namun setelahnya terjadi Penurunan pada bukan berikutnya. Pola tersebut berulang pada bulan 5/23 ke 6/23 terjadi kenaikan dan penurunan pada bulan berikutnya.

03

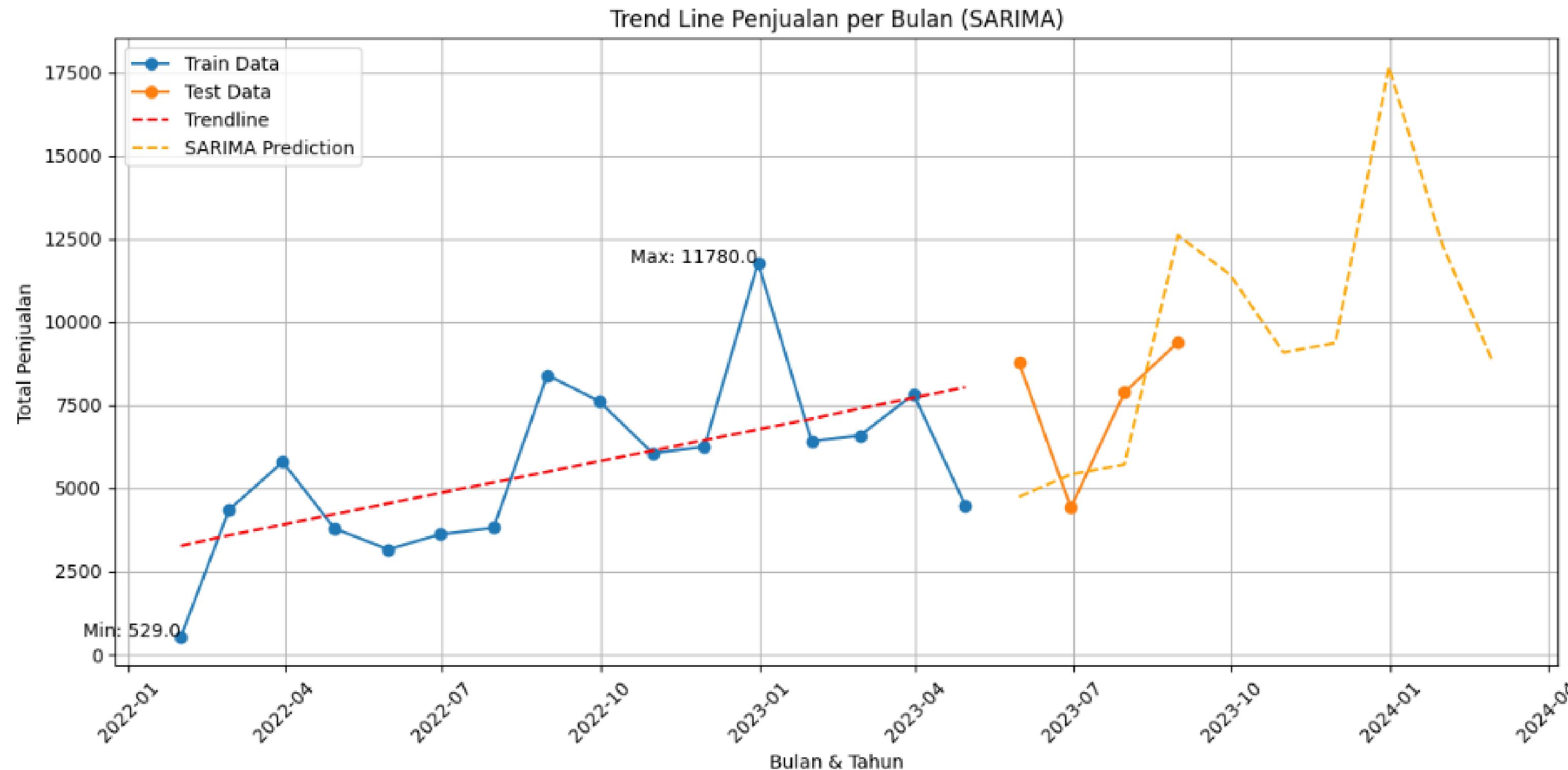
Bisa juga kita masukan dalam prediksi

Bisa juga kita masukan dalam prediksi “Bearish Head and Shoulders” dan “Bullish Flag Pattern”

BULLISH FLAG PATTERN & BEARISH HEAD AND SHOULDERS

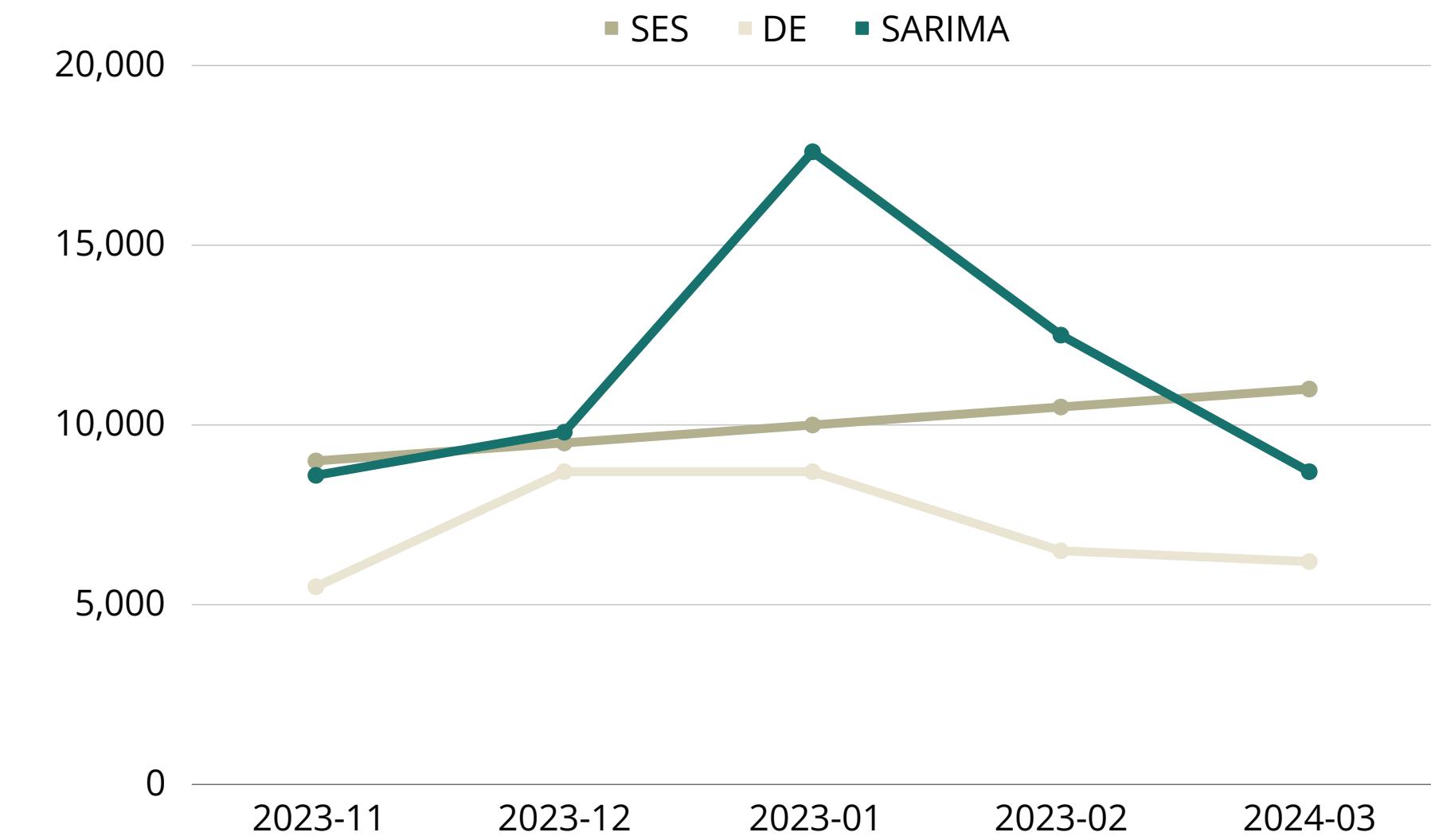


METODE TIME SERIES SARIMA (6 MONTH PREDICTION)



MODEL PREDICTION

Kesimpulan: Model SARIMA dengan parameter terbaik mampu memberikan prediksi yang cukup akurat untuk penjualan per bulan. Meskipun parameter yang digunakan terlihat sederhana (0, 0, 0) dan (0, 2, 0, 12), model ini mungkin cocok dengan pola penjualan yang ada. Dengan nilai AIC sekecil 2.0, model SARIMA menunjukkan kemampuan yang baik dalam menggambarkan tren dan musiman dari data penjualan. Rekomendasi: Perlu dilakukan evaluasi lebih lanjut untuk memahami apakah model ini memadai untuk memprediksi fluktuasi penjualan di masa mendatang. Disarankan untuk terus memonitor kinerja model dan melakukan penyesuaian jika diperlukan seiring berjalannya waktu.



- Tingkat terbaik hasil model prediction memberikan akurasi trend historical dengan tingkat tuning 96%
- Double Eksponensial memberikan prediksi yang cukup baik walau hasil predict sideways 83%
- Single Eksponensial memberikan prediksi yang lebih simple dengan membaca pola trend secara garis besar di 75% akurasi

SARAN



03

Fokus KPI kepenulis, untuk setiap penulis yang melakukan penerbitan buku dengan fokus topik yang sudah ditentukan bisa memberikan keuntungan insentif lebih ke penulis maupun ke penerbit

01

Tren Buku Sains dan Teknologi merupakan best seller. Kedepannya bisa menambahkan judul-judul buku yang mengikuti tren, seperti Machine Learning, Crypto currency, Metaverse, NFT, Web3, Ekonomi digital dll. Buku yang menjadi rekomendasi menjelang 2024 ini.
Link pendukung : [bitcoin halving](#)

02

- Buku tentang self-improvement dan pengembangan diri: Pemberdayaan diri, Pengelolaan stres, Produktivitas, Pengembangan karier, Hubungan interpersonal dll
Link Pendukung : [Buku self improvement](#)
- Buku tentang isu-isu aktual : Pemilu presiden 2024, Perang Rusia-Ukraina, Perubahan iklim, Kemiskinan, Ketidaksetaraan dll

Dari Analisa historical kita, memang teknologi and science menjadi topik yang paling tinggi secara grafik maka buku tentang tren terkini bisa menjadi fokus SKU deepublish untuk meningkatkan revenue

Link Pendukung : [Teknologi dan sains](#)



THANK YOU

● FOR YOUR NICE ATTENTION

November 2023