MethodsX
Open access
Article Template

Article information

Article title

Perbandingan Model ARIMAX Calendar Variation untuk Data Pencarian Google "Marjan"

Authors

Hammam Maulana Arijudin, Mohammad Imam Mahmudin, Muhammad Karunia Akbar, Irziqna Auliaurrachmah, Naifa Adila Bilqis, Nathania Rani Pujiati, Nurul Fadhila Yuni Setyawati, Khoirun Nisa

Affiliations

Departemen Statistika, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Corresponding author's email address and Twitter handle

5003231013@student.its.ac.id, 5003231031@student.its.ac.id, 5003231060@student.its.ac.id, 5003231061@student.its.ac.id, 5003231066@student.its.ac.id, 5003231073@student.its.ac.id, 5003231086@student.its.ac.id, 5003231111@student.its.ac.id

Keywords

Marjan; ARIMAX; *Calendar Variation*

Related research article

None

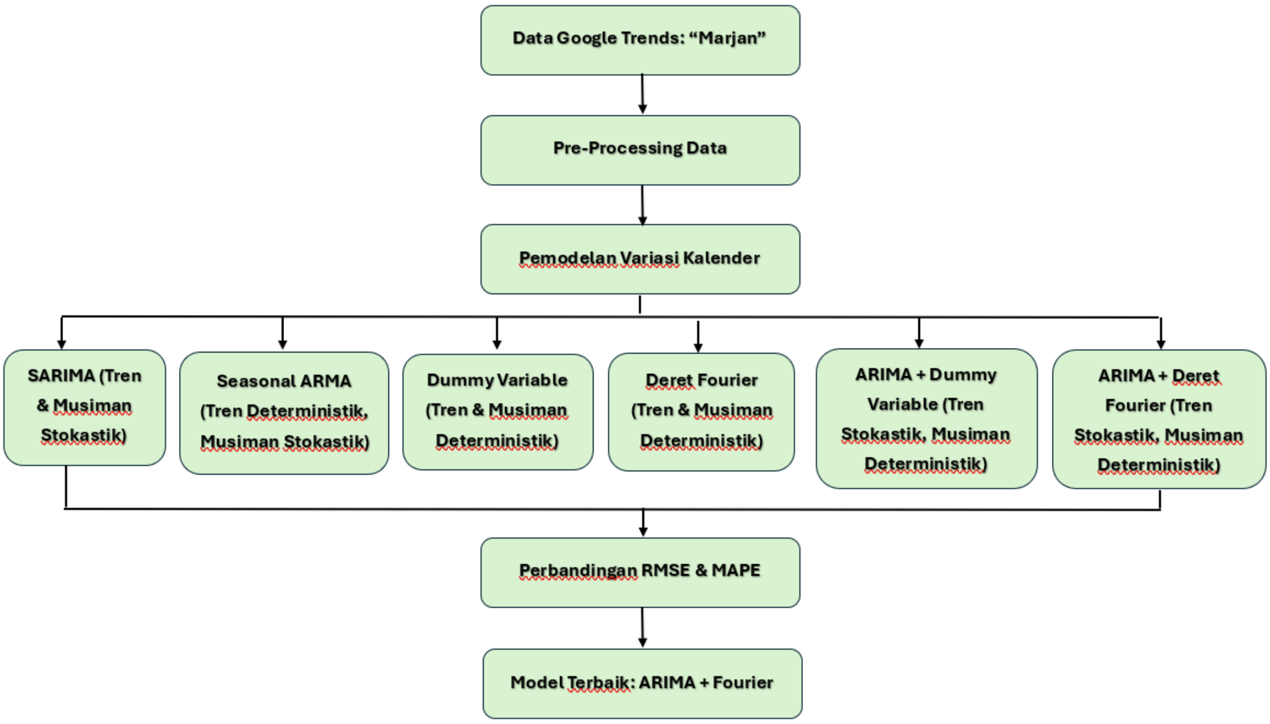
For a published article:

None

Abstract

Artikel penelitian ini menyajikan metode untuk meningkatkan peramalan deret waktu dengan memasukkan variasi kalender, yaitu pola musiman yang terjadi setiap tahun karena dipengaruhi hari raya/acara keagamaan atau budaya. Data yang digunakan berasal dari Google Trends untuk kata kunci "Marjan" di Indonesia, yang menunjukkan lonjakan musiman yang kuat menjelang Ramadan dan Idul Fitri. Untuk mengetahui jenis model yang cocok digunakan, dilakukan eksplorasi enam model peramalan berbasis ARIMA yang menggabungkan berbagai bentuk tren dan musiman, termasuk variabel *dummt* dan deret Fourier untuk efek kalender. Kinerja model dievaluasi menggunakan RMSE dan MAPE. Model terbaik yang didapat adalah model ARIMA(1,0,0)$(2,0,0)^{12}$ Deret Fourier, dengan MAPE 10,3363% yang menunjukkan model ini dapat menangkap efek variasi kalender yang terjadi.

Graphical abstract



Specifications table

Subject area	Mathematics and Statistics
More specific subject area	Time Series Analysis
Name of your method	ARIMAX Calendar Variation Model
Name and reference of original method	Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2008). Time Series Analysis: Forecasting and Control (edisi ke-4). Wiley. Tsay, R. S. (2010). Analysis of Financial Time Series (edisi ke-3). Wiley. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (edisi ke-2). OTexts.
Resource availability	Data deret waktu diambil dari Google Trends (https://trends.google.com) dengan kata kunci **"marjan"** untuk wilayah Indonesia, periode **Januari 2011 – Juni 2025**. Semua script pemodelan (R script) serta dataset yang telah diekstrak dilampirkan sebagai materi tambahan.

Background

Peramalan deret waktu merupakan salah satu teknik analisis yang penting dalam berbagai bidang, terutama ketika data menunjukkan adanya pola musiman atau tren dari waktu ke waktu. Salah satu tantangan utama dalam pemodelan deret waktu adalah bagaimana memisahkan dan mengidentifikasi komponen tren serta musiman, yang masing-masing dapat bersifat deterministik maupun stokastik. Ketepatan dalam menangkap kedua komponen ini sangat berpengaruh terhadap kinerja model peramalan.

Salah satu bentuk pola musiman yang menarik untuk dianalisis adalah variasi kalender, yaitu pola yang berulang secara konsisten pada periode-periode tertentu dalam satu tahun atau berdasarkan sistem penanggalan tertentu. Variasi ini dapat disebabkan oleh faktor budaya, sosial, maupun ekonomi yang berulang secara tahunan dan dapat memengaruhi nilai deret waktu secara signifikan.

Di Indonesia, kata kunci “Marjan” yang merupakan merek sirup populer menunjukkan pola musiman yang kuat setiap tahun menjelang bulan Ramadan dan Idul Fitri. Pola ini dapat diamati melalui data pencarian dari Google Trends, dan menjadikan data tersebut menarik untuk dianalisis menggunakan pendekatan variasi kalender dalam deret waktu.

Dalam studi ini, dilakukan analisis terhadap data Google Trends kata kunci "marjan" dari Januari 2011 hingga Juni 2025. Setelah melalui tahap praproses data, diterapkan enam model variasi kalender yang berbeda, masing-masing menggabungkan komponen tren dan musiman dalam bentuk deterministik maupun stokastik, yaitu:

1. Model Seasonal ARIMA (SARIMA): Tren dan musiman bersifat stokastik.
2. Model Seasonal ARMA dengan Tren Deterministik: Tren dimodelkan secara deterministik, sedangkan musiman tetap stokastik.
3. Model Dummy Variabel: Menggunakan variabel dummy untuk menangkap pengaruh musiman dan tren yang bersifat deterministic.
4. Model Deret Fourier: Menggunakan fungsi sinus dan cosinus dari deret Fourier untuk memodelkan pola musiman yang bersifat deterministik dan berulang secara periodik.
5. Model ARIMA dengan Musiman Dummy: Menggabungkan tren stokastik dengan komponen musiman deterministic melalui dummy variable
6. Model ARIMA dengan Musiman Fourier: Menggabungkan tren stokastik dengan komponen musiman deterministic melalui deret Fourier

Tujuan dari studi ini adalah untuk mengevaluasi dan membandingkan kinerja keenam model tersebut berdasarkan nilai galat peramalan, seperti RMSE dan MAPE, guna menentukan model mana yang paling efektif dalam menangkap pola musiman berbasis kalender serta menghasilkan peramalan yang akurat di luar sampel (*out-sample*).

Method details

1. Pre-Processing Data

Analisis ini menggunakan data indeks pencarian bulanan keyword “Marjan” dari Google Trends, periode Januari 2011 hingga Juni 2025. Data dibagi menjadi data in-sample (Januari 2011 – Desember 2022) dan data out-sample (Januari 2023 – Juni 2025). Untuk memeriksa stasioneritas dalam varians, dilakukan perhitungan Box-Cox dan didapat hasil $\lambda = 0$ (Gambar 1). Maka, data tidak stasioner dalam varians dan ditransformasi dengan $\ln{(Y)}$. Selain itu, dilakukan Augmented Dickey-Fuller (ADF) test untuk menguji stasioneritas dalam rata-rata dan didapat nilai p-value 0.282 (Gambar 2) yang berarti data belum stasioner dalam rata-rata dan terdapat pola tren pada data.

Method

Seasonal period	12
Select optimal λ from interval	[-1; 1]
Optimal λ	-0,256830
Rounded optimal λ	0
Transformed series = ln(Y)	

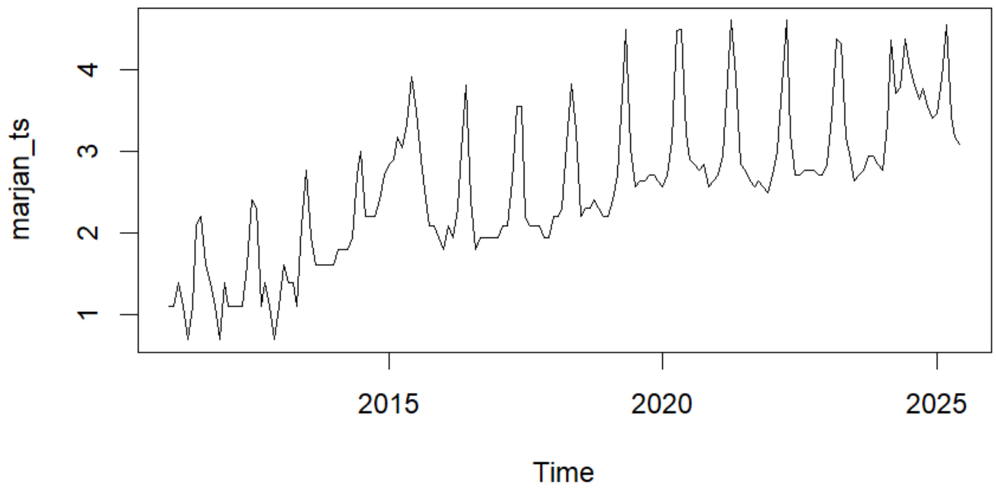
Gambar 1. Hasil Perhitungan Box-Cox dengan Minitab

Augmented Dickey-Fuller Test

Null hypothesis: Data are non-stationary
Alternative hypothesis: Data are stationary

Test Statistic	P-Value	Recommendation
-2,01153	0,282	Test statistic > critical value of -2,87967. Significance level = 0,05 Fail to reject null hypothesis. Consider differencing to make data stationary.

Gambar 2. Hasil Augmented Dickey-Fuller Test



Gambar 3. Time Series Plot Data Setelah Ditransformasi

2. Model Variasi Kalender dengan Efek Tren dan Musiman Stokastik (Seasonal ARIMA)

Bentuk umum model variasi kalender dengan efek tren & musiman stokastik dengan Seasonal ARIMA (p,d,q) (P,D,Q)[S] dapat dituliskan sebagai berikut:

$$y_t = \sum_{j=1}^{J} \beta_j D_{j,t} + \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D}\epsilon_t$$

Keterangan:
y_t: data penjualan marjan pada bulan ke-t
β_j: Parameter koefisien efek kalender
$D_{j,t}$: Dummy variable untuk hari raya idul fitri (bernilai 0 jika tidak terjadi idul fitri, bernilai 1 jika terjadi idul fitri pada bulan tersebut)
ϵ_t: Error pada bulan ke-t
S: Panjang periode musim
Estimasi parameter

Model diestimasi menggunaka ARIMA$(0,1,2)(2,0,1)^{12}$, menggunakan 1 *dummy* kalendar dengan estimasi sebagai berikut:

$$ln(y_t) = -0.0927\ D_t + \frac{(1-0.1851B-0.098B^2)(1-0.7211B^S)}{(1+1.6465B^{12}-0.7783B^{24})(1-B)^1}\epsilon_t$$

3. Model Variasi Kalender dengan Trend Deterministik dan Musiman Stokastik (Seasonal ARMA)

Model deret waktu yang mempertimbangkan variasi kalender dengan tren deterministik dan komponen musiman stokastik (Seasonal ARMA) dapat dituliskan dalam bentuk umum sebagai berikut:

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{J} \beta_j D_{j,t} + \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B)^D}\epsilon_t$$

Keterangan:

Y_t: Variabel dependen pada waktu t

β_0: Intersep (konstanta)

β_1: Koefisien tren linear deterministik

β_j: Koefisien efek kalender

$D_{j,t}$: *Dummy variable* untuk efek kalender (misal: 1 jika ada hari raya di bulan t, 0 lainnya)

S: Panjang periode musim

Berdasarkan analisis yang telah dilakukan, model *Seasonal* ARIMA (SARIMA) dengan pendekatan variasi kalender, tren deterministik, dan musiman stokastik berhasil dibangun untuk memprediksi penjualan Marjan. Model terpilih adalah ARIMA(1,0,0)(2,0,1)^{12} yang diidentifikasi secara otomatis menggunakan fungsi auto.arima() dengan estimasi sebagai berikut:

$$\ln(Y_t) = 1.3938 + 0.0127t - 0.0987D_{j,t} + \frac{(1-0.7275B^{12})}{(1-0.7764B)(1-1.621B^{12}+0.7721B^{24})}\epsilon_t$$

4. Model Variasi Kalender dengan Trend Deterministik dan Musiman Deterministik (Dummy Variable)

Model umum untuk model variasi kalender dengan trend deterministik dan musiman deterministik (*dummy variable*) dalam analisis deret waktu biasanya dinyatakan alam bentuk regresi time series, Model umum tersebut dapat dituliskan sebagai berikut.

$$Y_t = \beta_0 + \beta_1 t + \sum_{i=1}^{s-1} \gamma_i M_{i,t} + \sum_{j=1}^{k} \delta_j D_{j,t} + \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)}\epsilon_t$$

Keterangan:

Y_t: Variabel dependen pada waktu ke-t

β_0: Intercept (konstanta)

$\beta_1 t$: Komponen trend deterministik (linear terhadap waktu)

$M_{i,t}$: Variabel dummy musiman ($i = 1$: Januari, $i = 2$: Februari, dst.)

γ_i: Parameter efek musiman

$D_{j,t}$: Variabel dummy untuk Hari Raya Idul Fitri (bernilai 0 jika tidak terjadi idulfitri, bernilai 1 jika terjadi idulfitri pada bulan tersebut)

δ_j: Parameter efek variasi kalender

ε_t: Error pada waktu ke-t

Berikut merupakan hasil estimasi parameter dengan menggunakan model ARIMA(1,0,0)(2,0,0)^{12} dengan trend deterministik, variabel dummy untuk bulan dan 1 dummy kalender.

$$\ln(Yt) = 0.8879 + 0.0135t + 0.1941M_1 + 0.3418M_2 + 0.6185M_3 + 0.9913M_4 + 0.8533M_5 + 0.7874M_6 + 0.7041M_7 + 0.5363M_8 + 0.2952M_9 + 0.2272M_{10} + 0.108M_{10} - 0.1149D_t + \frac{1}{(1-0.7497B)(1-0.9444B^{12}+0.2732B^{24})}\epsilon_t$$

5. Model Variasi Kalender dengan Trend Deterministik dan Musiman Deterministik (Deret Fourier)

Bentuk umum model ARIMAX variasi kalender dengan tren deterministik dan musiman deterministik (deret fourier) adalah

$$y_t = \mu + \delta_1 t + \sum_{k=1}^{K}\left[\alpha_k \cos\left(\frac{2\pi kt}{S}\right) + \gamma_k \sin\left(\frac{2\pi kt}{S}\right)\right] + \sum_{j=1}^{J}\beta_j D_{j,t} + \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D}\epsilon_t$$

μ: intercept atau konstanta
δ_1: koefisien tren
$D_{j,t}$: Dummy variable untuk hari raya idulfitri (bernilai 0 jika tidak terjadi idulfitri, bernilai 1 jika terjadi idulfitri pada bulan tersebut)
K: harmonik untuk deret fourier (semakin tinggi, pola musiman semakin kompleks)
S: Panjang periode musim

Pada kasus ini, dipilih $K = 1$ karena pola musiman terlihat sederhana dan untuk menghindari over-fitting. Orde model SARIMAX dipilih secara otomatis menggunakan fungsi auto.arima() dari package forecast di R yang didasarkan pada nilai AIC terkecil. Didapatkan hasil model ARIMA(1,0,0)(2,0,0)^{12}, t, deret fourier, D_t dengan estimasi parameter sebagai berikut.

$$\ln y_t = 1.3638 + 0.0135\,t - 0.3933\cos\left(\frac{2\pi t}{12}\right) + 0.1548\sin\left(\frac{2\pi t}{12}\right) - 0.1237\,D_t + \frac{1}{(1-0.7841B)(1-0.9555B^{12}+0.2811B^{24})}\epsilon_t$$

6. Model Variasi Kalender dengan Trend Stokastik (ARIMA) dan Musiman Deterministik (Dummy Variable)

Bentuk umum model ARIMAX variasi kalender dengan tren stokastik dan musiman deterministik (dummy variable) adalah

$$y_t = \mu + \sum_{s=1}^{S-1}\beta_s M_s + \sum_{j=1}^{J}\alpha_j D_{j,t} + \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D}\epsilon_t$$

μ: intercept atau konstanta
M_s: dummy variable untuk musim
$D_{j,t}$: ummy variable untuk hari raya idulfitri (bernilai 0 jika tidak terjadi idulfitri, bernilai 1 jika terjadi idulfitri pada bulan tersebut)
S: Panjang periode musim

Setelah dilakukan identifikasi model dan estimasi parameter secara otomatis dengan fungsi auto.arima() di R, didapatkan model ARIMA(1,1,1)(1,0,2)^{12} sebagai berikut.

$$\begin{aligned} y_t = {} & 0.2138\,M_1 + 0.3431\,M_2 + 0.6466\,M_3 + 1.002\,M_4 + 0.8068\,M_5 + 0.7293\,M_6 + 0.697\,M_7 + 0.5586\,M_8 \\ & + 0.3073\,M_9 + 0.2429\,M_{10} + 0.1229\,M_{11} - 0.0836\,D_t \\ & + \frac{(1+0.9745B)(1-0.2779B^{12}-0.1972B^{24})}{(1-0.7722B)(1-0.5934B^{12})(1-B)}\epsilon_t \end{aligned}$$

7. Model Variasi Kalender dengan Trend Stokastik (ARIMA) dan Musiman Deterministik (Deret Fourier)

Model *trend* stokastik ARIMA, musiman deterministik dengan menggunakan deret Fourier, dan variabel *dummy* sebagai efek variasi kalender dikategorikan sebagai model ARIMAX Fourier *with Calendar Effects.* Model ini dirancang untuk menangkap pola tren acak dalam data deret waktu melalui komponen ARIMA, serta pola musiman deterministik melalui pendekatan deret Fourier, serta efek khusus yang tidak berulang secara reguler ditangkap melalui variabel *dummy* kalender.

Secara matematis, model dituliskan sebagai berikut.

$$y_t = \sum_{k=1}^{K}\left[\alpha_k \cos\left(\frac{2\pi kt}{S}\right) + \gamma_k \sin\left(\frac{2\pi kt}{S}\right)\right] + \sum_{j=1}^{J}\beta_j D_{j,t} + \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D}\epsilon_t$$

Keterangan:
$D_{j,t}$: *Dummy* variable untuk hari raya idulfitri (bernilai 0 jika tidak terjadi idulfitri, bernilai 1 jika terjadi idulfitri pada bulan tersebut)

K: harmonik untuk deret fourier (semakin tinggi, pola musiman semakin kompleks)

S: Panjang periode musim

Estimasi parameter

Model diestimasi menggunaka ARIMA$(1,1,1)(2,0,0)^{12}$, menggunakan satu pasang Fourier *term* (K=1), dan 1 *dummy* kalender dengan estimasi sebagai berikut.

$$\ln y_t = -0.3866 \cos\left(\frac{2\pi t}{12}\right) + 0.1498 \sin\left(\frac{2\pi t}{12}\right) - 0.127 D_t + \frac{(1-0.9746B)(1)}{(1-0.7780B)(1-0.9774B^{12})(1+0.2811B^{24})(1-B)^1}\varepsilon_t$$

Method validation

Perbandingan RMSE dan MAPE Out-Sample Antarmodel

Untuk mengevaluasi kinerja model variasi kalender yang dibuat, root mean squared error (RMSE) dan mean absolute percentage error (MAPE) dipilih sebagai metrik evaluasi. RMSE dan MAPE dari out-sample didefinisikan sebagai berikut.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_t - \hat{y}_t)^2}{n}} \quad \text{dan} \quad \text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100\%$$

Hasil RMSE dan MAPE out-sample dari masing-masing model disajikan dalam tabel berikut.

Tabel 1. RMSE dan MAPE Out-Sample

Model	RMSE	MAPE
ARIMA(0,1,2)(2,0,1)^{12}, $D_{j,t}$	0.7133	14.7759%
ARIMA(1,0,0)(2,0,1)^{12}, t, $D_{j,t}$	0.5268	10.4593%
ARIMA(1,0,0)(2,0,0)^{12}, t, Dummy Variable, $D_{j,t}$	0.4462	10.5189%
ARIMA(1,0,0)(2,0,0)^{12}, t, Deret Fourier, D_t	0.4352	10.3363%
ARIMA(1,1,1)(1,0,2)^{12}, Dummy Variable, D_t	0.5908	12.1271%
ARIMA(1,1,1)(2,0,0)^{12}, $D_{j,t}$, Deret Fourier	0.5955	12.1158%

Berdasarkan hasil tersebut, dengan mempertimbangkan MAPE terkecil, maka model ARIMA(1,0,0)(2,0,0)^{12}, t, Deret Fourier, D_t adalah model yang paling cocok untuk data trend Marjan. Nilai MAPE sebesar 10.3363% ini menunjukkan bahwa kesalahan prediksi yang dihasilkan model ini relatif kecil dibandingkan model lainnya.

Limitations

Model ARIMA(1,0,0)(2,0,0)^{12} dengan tren dan komponen Fourier memberikan MAPE terkecil, namun metode ini memiliki beberapa keterbatasan. Salah satunya, Fourier yang dapat menyebabkan overfitting jika digunakan berlebihan atau jika data deret waktu relatif pendek. Model ini menjadi sensitif terhadap kualitas data. Selain itu, model ini tidak bersifat adaptif secara *real-time,* sehingga perlu dilakukan pemantauan dan pelatihan secara berkala agar akurasinya tetap terjaga.

Credit author statement

Hammam Maulana Arijudin : menyusun dan menganalisis model 1
Mohammad Imam Mahmudin : menyusun dan menganalisis model 4
Muhammad Karunia Akbar : menyusun dan menganalisis model 5
Irziqna Auliaurrachmah : menyusun dan menganalisis model 6
Naifa Adila Bilqis : menyusun dan menganalisis model 2
Nathania Rani Pujiati : menyusun laporan

Nurul Fadhila Yuni Setyawati : menyusun laporan
Khoirun Nisa : menyusun dan menganalisis model 3

References

Azizah, N. (2017). *Penerapan metode autoregressive integrated moving average with exogenous variables (ARIMAX) berdasarkan variasi kalender Hijriyah pada peramalan penjualan busana Muslim* [Skripsi sarjana, Institut Teknologi Sepuluh Nopember]. Institut Teknologi Sepuluh Nopember.

Cryer, J. D., & Chan, K.-S. (2008). *Time series analysis: With applications in R* (2nd ed.). Springer Science+Business Media.

Hanke, J. E., & Wichern, D. W. (2014). *Business forecasting* (9th ed.). Pearson Education Limited.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (edisi ke-3). OTexts. Diakses pada 2 Juli 2025 dari https://OTexts.com/fpp3