

## Application of the Autoregressive Integrated Moving Average Exogenous (ARIMAX) with Calendar Variation Effect Method for Forecasting Chocolate Data in Indonesia and the United States

### Penerapan Metode *Autoregressive Integrated Moving Average Exogenous* (ARIMAX) with *Calendar Variation Effect* untuk Peramalan Data Cokelat di Indonesia dan Amerika Serikat

\*Andy Rezky Pratama Syam

#### Abstract

Forecasting chocolate consumption is required by producers in preparing the amount of production each month. The tradition of Valentine, Christmas and Eid al-Fitr which are closely related to chocolate makes it impossible to predict chocolate by using the Classical Time Series method. Especially for Eid al-Fitr, the determination follows the Hijri calendar and each year advances 10 days on the Masehi calendar, so that every three years Eid al-Fitr will occur in a different month. Based on this, the chocolate forecasting will show a variation calendar effect. The method used in modeling and forecasting chocolate in Indonesia and the United States is the ARIMAX (Autoregressive Integrated Moving Average Exogenous) method with Calendar Variation effect. As a comparison, modeling and forecasting are also carried out using the Naïve Trend Linear, Naïve Trend Exponential, Double Exponential Smoothing, Time Series Regression, and ARIMA methods. The ARIMAX method with Calendar Variation Effect produces a very precise MAPE value in predicting chocolate data in Indonesia and the United States. The resulting MAPE value is below 10 percent, so it can be concluded that this method has a very good ability in forecasting.

**Keywords:** *Chocolate, ARIMAX, MAPE, Calendar Variation*

#### Abstrak

Peramalan konsumsi cokelat diperlukan oleh produsen dalam mempersiapkan jumlah produksi setiap bulannya. Adanya tradisi Valentine, hari Natal dan hari raya idul Fitri yang sangat berkaitan erat dengan cokelat membuat peramalan cokelat tidak dapat dilakukan dengan menggunakan metode Time Series Klasik. Khusus untuk Hari Raya Idul Fitri, penetapannya mengikuti kalender Hijriyah dan setiap tahunnya maju 10 hari pada kalender Masehi, sehingga setiap tiga tahun Hari Raya Idul Fitri akan terjadi di bulan yang berbeda. Berdasarkan hal tersebut, maka dalam peramalan cokelat akan muncul efek kalender variasi. Metode yang digunakan dalam pemodelan dan peramalan cokelat di Indonesia dan Amerika Serikat adalah metode ARIMAX (Autoregressive Integrated Moving Average Exogenous) with Calendar Variation effect. Sebagai pembanding, dilakukan juga pemodelan dan peramalan dengan metode Naïve Trend Linear, Naïve Trend Exponensial, Double Exponential

\*Badan Pusat Statistik Kabupaten Bulukumba

\*Email address: [andy.pratama@bps.go.id](mailto:andy.pratama@bps.go.id)



Smoothing, Time Series Regression, dan ARIMA. Metode ARIMAX with Calendar Variation Effect menghasilkan nilai MAPE yang sangat presisi dalam melakukan peramalan terhadap data coklat di Indonesia dan Amerika Serikat. Nilai MAPE yang dihasilkan berada di bawah 10 persen, sehingga dapat disimpulkan metode tersebut memiliki kemampuan yang sangat baik dalam melakukan peramalan.

**Kata Kunci :** Cokelat, ARIMAX, MAPE, Kalender Variasi

## I. PENDAHULUAN

Cokelat adalah salah satu produk olahan kakao yang banyak digemari masyarakat karena mempunyai citarasa yang khas. Cokelat merupakan kategori makanan yang mudah dicerna oleh tubuh dan mengandung banyak vitamin seperti vitamin A1, B1, B2, C, D, dan E serta beberapa mineral seperti fosfor, magnesium, zat besi, zinc, dan juga tembaga [7]. Berdasarkan data dari Kementrian Pertanian Republik Indonesia, terjadi peningkatan produksi coklat yang sangat signifikan di Indonesia, dimana produksi coklat tahun 2017 sebesar 590.684 ton menjadi 734.796 ton pada tahun 2019. Jika dirinci berdasarkan provinsi, pada tahun 2019 provinsi Sulawesi Tengah merupakan produsen coklat terbesar di Indonesia, dimana hasil produksinya mencapai 128.154 ton. Dalam kurun waktu dua tahun, terjadi peningkatan produksi coklat sebesar 178.542 ton. Hal ini sejalan dengan permintaan coklat yang semakin banyak baik dari dalam negeri maupun dari luar negeri.

Peramalan konsumsi coklat diperlukan oleh produsen dalam mempersiapkan jumlah produksi setiap bulannya. Adanya tradisi Valentine dan hari raya idul Fitri yang sangat berkaitan erat dengan coklat membuat peramalan coklat tidak dapat dilakukan dengan menggunakan metode Time Series Klasik. Khusus untuk Hari Raya Idul Fitri, penetapannya mengikuti kalender Hijriyah dan setiap tahunnya maju 10 hari pada kalender Masehi, sehingga setiap tiga tahun Hari Raya Idul Fitri akan terjadi di bulan yang berbeda. Berdasarkan hal tersebut, maka dalam peramalan coklat akan muncul efek kalender variasi.

Data Google Trend dapat digunakan untuk melakukan pemodelan dan peramalan terhadap suatu topik tertentu. Oleh karena itu, untuk dapat melakukan pemodelan dan peramalan terhadap Cokelat, dilakukan pencarian data melalui google trend terhadap kata "Chocolate" di Amerika Serikat dan kata "Cokelat" di Indonesia yang diprediksi memiliki pola musiman dan efek kalender variasi. Apabila dalam melakukan peramalan menggunakan model ARIMA (Autoregressive Integrated Moving Average), variasi kalender dapat diidentifikasi sebagai outlier. Oleh karena itu, diperlukan penambahan variabel dummy untuk mengatasi adanya variasi kalender. Model yang dapat digunakan adalah model ARIMAX (*Autoregressive Integrated Moving Average Exogenous*).

Lee [3] meramalkan banyak penjualan baju muslim laki-laki yang selalu meningkat tajam saat Hari Raya Idul Fitri akan tiba. Model ARIMAX merupakan model yang terbaik untuk meramalkan banyak penjualan baju muslim laki-laki tersebut, dibandingkan dengan model-model lainnya. Anggraeni [1] meramalkan penjualan baju muslim anak-anak Habibah Busana yang selalu meningkat menjelang libur lebaran dan menurun di bulan berikutnya. Hasilnya menunjukkan bahwa model ARIMAX lebih baik daripada ARIMA.

Berdasarkan latar belakang tersebut, akan dilakukan pemodelan dan peramalan terhadap Cokelat di negara Indonesia dan Amerika Serikat. Untuk mendapatkan model yang terbaik, akan digunakan beberapa jenis metode yaitu ARIMA (*Autoregressive Integrated Moving Average*), TSR (*Time Series Regression*), dan ARIMAX (*Autoregressive Integrated Moving Average Exogenous*) with *Calendar Variation effect*.

Tujuan dari penelitian ini adalah untuk mendapatkan hasil ramalan yang tepat terhadap data coklat di Indonesia dan Amerika dari google trend yang mengandung pola musiman dan

efek kalender variasi. Selain itu, akan dilakukan perbandingan nilai MAPE (*Mean Absolute Percentage Error*) dari setiap metode untuk mendapatkan metode terbaik.

## II. METODOLOGI

Untuk menjawab tujuan dari penelitian, dilakukan simulasi ramalan dengan menggunakan beberapa jenis metode statistik yang relevan terhadap data coklat di Indonesia dan Amerika. Metode analisis yang digunakan merupakan metode *Time Series* yaitu *Naïve Trend Linear*, *Naïve Trend Seasonal*, *Double Exponential Smoothing*, *Time Series Regression*, *ARIMA*, dan *ARIMAX with Calendar Variation*. Data coklat di Indonesia dan Amerika terlebih dahulu dibagi menjadi data *Training* dan *Testing*, dimana data *Training* akan digunakan untuk memperoleh model *Time Series* terbaik sedangkan data *Testing* akan digunakan untuk *forecasting*. Metode statistik yang telah ditetapkan akan diterapkan terhadap data *training* dan akan dibandingkan nilai MAPE dari setiap metode. Metode statistik dengan nilai MAPE yang terkecil merupakan metode statistik dengan peramalan yang paling akurat. Sebagai bentuk *forecasting*, akan dilakukan peramalan untuk 1 bulan kedepan, 3 bulan kedepan, 6 bulan kedepan, 1 tahun kedepan dan 2 tahun kedepan.

### 2.1 Data Time Series

Data Time series atau Deret waktu adalah himpunan observasi data terurut dalam waktu [8]. Metode time series adalah metode peramalan dengan menggunakan analisa pola hubungan antara variabel yang akan dipikirkan dengan variabel waktu. Peramalan suatu data time series perlu memperhatikan tipe atau pola data. Secara umum terdapat empat macam pola data time series, yaitu horizontal, trend, musiman, dan siklis [8]. Pola horizontal merupakan kejadian yang tidak terduga dan bersifat acak, tetapi kemunculannya dapat memengaruhi fluktuasi data time series. Pola trend merupakan kecenderungan arah data dalam jangka panjang, dapat berupa kenaikan maupun penurunan. Pola musiman merupakan fluktuasi dari data yang terjadi secara periodik dalam kurun waktu satu tahun, seperti triwulan, kuartalan, bulanan, mingguan, atau harian. Sedangkan pola siklis merupakan fluktuasi dari data untuk waktu yang lebih dari satu tahun.

### 2.2 Naïve Method

Metode peramalan *naïve* merupakan metode peramalan berdasarkan pengamatan pola data sebelumnya, sehingga metode ini mengasumsikan bahwa data masa lalu sebagai indikator peramalan terbaik di masa depan. Jika pola data musiman maka metode *naïve* musiman memodelkan data masa depan berdasarkan data musiman di masa lalu [11].

Dalam penelitian ini digunakan 2 jenis metode peramalan *Naïve*, yaitu :

- *Naïve Trend Linear*

Bentuk model *Naïve* apabila data mengandung trend linear adalah :

$$\hat{y}_{t+1} = y_t + (y_t - y_{t-1})$$

(2.1)

- *Naïve Trend Exponensial*

Bentuk model *Naïve* apabila data mengandung trend Exponensial adalah :

$$\hat{y}_{t+1} = y_t \frac{y_t}{y_{t-1}}$$

(2.2)

Jika data time series memuat unsur seasonal,  $s$ , maka peramalan pada metode *Naïve* didasarkan pada pengamatan pada periode yang sama pada seasonal sebelumnya. Jika memuat

pola double seasonal, dimana  $s_1 < s_2$ , maka peramalan pada metode Naïve didasarkan pada periode seasonal terpanjang, yaitu  $s_2$ . Secara matematis, peramalan menggunakan metode seasonal Naïve dan double seasonal Naïve dapat dituliskan sebagai berikut [10].

$$\hat{Y}_t(h) = Y_{t+h-s} \quad (2.3)$$

$$\hat{Y}_t(h) = Y_{t+h-s_2} \quad (2.4)$$

### 2.3 Double Exponential Smoothing

Metode *double exponential smoothing* dibagi menjadi *double exponential smoothing* satu parameter dari *Brown* dan dua parameter dari *Holt*. Metode *double exponential smoothing* satu parameter dari *Brown* dikembangkan untuk mengatasi perbedaan yang muncul antara data aktual dan nilai peramalan apabila ada trend pada pola datanya [10]. Adapun rumus yang digunakan dalam metode *double exponential smoothing* satu parameter dari *Brown* dapat ditentukan berdasarkan persamaan :

- Nilai *Single Exponential Smoothing*

$$S'_t = \alpha X_t + (1 - \alpha)S'_{t-1} \quad (2.5)$$

- Nilai *Double Exponential Smoothing*

$$S''_t = \alpha S_t + (1 - \alpha)S''_{t-1} \quad (2.6)$$

- Nilai konstanta

$$a_t = S'_t + (S'_t - S''_t) = 2S'_t - S''_t \quad (2.7)$$

- Nilai *Trend*

$$b_t = \frac{\alpha}{1 - \alpha}(S'_t - S''_t) \quad (2.8)$$

- Nilai peramalan

$$F_{t+m} = a_t + b_t m \quad (2.9)$$

Dimana

$S'_t$  : Nilai *single exponential smoothing* periode ke  $t$

$\alpha$  : Parameter *exponential smoothing* ( $0 < \alpha < 1$ )

$X_t$  : Data aktual pada periode  $t$

$S'_{t-1}$  : Nilai *single exponential smoothing* periode ke  $t-1$

$S''_t$  : Nilai *double exponential smoothing* periode ke  $t$

$S''_{t-1}$  : Nilai *double exponential smoothing* periode ke  $t-1$

$a_t$  : Nilai konstanta pada periode ke- $t$

$b_t$  : Nilai trend pada periode ke- $t$

$m$  : Periode ke depan yang akan diramalkan

$F_{t+m}$  : Nilai peramalan untuk  $m$  periode ke depan

Metode *double eksponential smoothing linier* dari Holt memuluskan nilai trend dengan parameter yang berbeda dari parameter yang digunakan pada pemulusan data yang asli. Adapun rumus yang digunakan dalam metode *double exponential smoothing* dua parameter dari Holt dapat dilihat pada persamaan :

- Nilai *Single Exponential Smoothing*

$$S'_t = \alpha X_t + (1 - \alpha)(S'_{t-1} + b_{t-1}) \quad (2.10)$$

- Nilai pemulusan *Trend*

$$b_t = \gamma(S'_t - S'_{t-1}) + (1 - \gamma)b_{t-1} \quad (2.11)$$

- Nilai Ramalan

$$F_{t+m} = S'_t + b_t m \quad (2.12)$$

Dimana :

- $S'_t$  : Nilai *single exponential smoothing* periode ke  $t$   
 $S'_{t-1}$  : Nilai *single exponential smoothing* periode ke  $t-1$   
 $\alpha$  : Nilai parameter exponential smoothing ( $0 < \alpha < 1$ )  
 $X_t$  : Data aktual pada periode ke- $t$   
 $b_t$  : Pemulusan trend pada periode ke  $t$   
 $b_{t-1}$  : Pemulusan trend pada periode ke  $t-1$   
 $\gamma$  : Nilai parameter pemulusan trend ( $0 < \gamma < 1$ )  
 $m$  : Periode ke depan yang akan diramalkan  
 $F_{t+m}$  : Nilai peramalan untuk  $(t+m)$  periode ke depan

Proses inialisasi untuk pemulusan eksponensial ganda dari Holt memerlukan dua taksiran, yaitu  $S_t$  dan  $b_t$ . proses inialisasi, diawali dengan memilih  $S_1 = X_1$  sedangkan untuk taksiran trend didapat dari rumus  $b_1 = X_2 - X_1$ . Adapun konstanta pemulusan  $\alpha$  dan  $\gamma$  dalam metode peramalan *double exponential smoothing* berperan sebagai faktor pembobotan. Nilai  $\alpha$  dan  $\gamma$  bernilai dekat dengan satu, ramalan terbaru akan menyertakan penyesuaian yang besar untuk setiap kesalahan yang terjadi pada ramalan sebelumnya. Sebaliknya jika  $\alpha$  dan  $\gamma$  dekat dengan nol, maka akan dihasilkan ramalan terbaru yang akan sangat mirip dengan nilai yang lama.

#### 2.4 Model ARIMA Box-Jenkins

Autoregressive Integrated Moving Average (ARIMA) dikelompokkan menjadi tiga, yaitu Autoregressive (AR), Moving Average (MA), dan model campuran dengan karakteristik keduanya, yakni Autoregressive Moving Average (ARMA). Model ARMA yang melalui proses integrasi, dinyatakan sebagai ARIMA[9].

Prosedur Box dan Jenkins [9] meliputi empat tahapan untuk membentuk model ARIMA, yaitu (1) identifikasi model, (2) penaksiran parameter, (3) pemeriksaan diagnostik, dan (4) peramalan. Secara model umum ARIMA ( $p, d, q$ ) dapat ditulis dalam bentuk

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)a_t \quad (2.13)$$

Jika terdapat efek musiman atau non-musiman maka model ARIMA yang terbentuk yaitu model multiplikatif ARIMA ( $p, d, q$ )( $P, D, Q$ )s jika ditulis secara matematis dapat dituliskan sebagai berikut

$$\Phi_P(B^S)\phi_p(B)(1-B)^d(1-B) \quad (2.14)$$

dengan :

$$\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p) \quad (2.15)$$

$$\theta_q(B) = (1 - \theta_1 B^s - \dots - \theta_q B^q) \quad (2.16)$$

$$\Phi_P(B) = (1 - \Phi_1 B^S - \dots - \Phi_P B^{Ps}) \quad (2.17)$$

$$\Theta_Q(B) = (1 - \Theta_1 B^S - \dots - \Theta_Q B^{Qs}) \quad (2.18)$$

dengan

$B$  = Operator *backward shift*

$s$  = Periode musiman

$d$  dan  $D$  = orde *differencing* non-musiman dan musiman

Sedangkan Seasonal Autoregressive Integrated Moving Average (SARIMA) adalah pengembangan dari model ARIMA pada data runtun waktu yang memiliki pola musiman. Notasi SARIMA adalah: SARIMA ( $p, d, q$ ) ( $P, D, Q$ )<sup>S</sup>. Dengan  $p, d, q$ : bagian yang tidak musiman dari model; ( $P, D, Q$ )<sup>S</sup>: bagian musiman dari model; S: jumlah periode per musim. Rumus umum SARIMA dapat dituliskan:  $\phi_p B^s \Phi_P(B)(1-B)^d(1-B^S)^D Z_t = \theta_q(B)\Theta_Q(B^S)a_t$ . Dengan:  $\phi_p(B)$  = AR *non-seasonal*;  $(1-B)^d$  = *differencing non-seasonal*;  $(1-B^S)^D$  = *differencing seasonal*;  $\theta_q(B)$  = MA

*non-seasonal*;  $\Theta_q(B^S) = MA$  *seasonal*. Untuk proses SARIMA yang umum, fungsi peramalan jangka panjang bergantung pada pola musiman individual paling akhir dalam data percobaan, dan peramalan terdiri atas penambahan periodik pola dan komponen modifikasi aditif (pertambahan) yang cenderung sama dengan nol secara eksponensial secara cepat karena horizon bertambah besar.

## 2.5 Model ARIMAX

Pemodelan deret waktu dengan menambahkan peubah yang dianggap memiliki pengaruh yang signifikan terhadap data sering kali dilakukan untuk menambah akurasi peramalan. Pemodelan Autoregressive Integrated Moving Average Exogenous (ARIMAX) merupakan modifikasi dari model ARIMA dengan tambahan peubah yang memiliki pengaruh signifikan. Faktor-faktor yang mempengaruhi peubah repon  $Y$  pada waktu  $t$  tidak hanya berasal dari fungsi peubah  $Y$  dalam waktu  $t$ , tetapi juga peubah lain yakni prediktor pada waktu ke  $t$ .

Model ARIMAX terbagi menjadi dua, yaitu model ARIMAX dengan tren stokastik dan model ARIMAX dengan tren deterministik. Tren pada data deret waktu yang mudah diprediksi dan tidak terlalu bervariasi disebut sebagai tren deterministik. Tren stokastik adalah tren yang sulit diprediksi dan cukup bervariasi. Model tren deterministik merupakan model tren stokastik dengan komponen deterministik di dalamnya, yaitu  $\gamma$  yang mewakili periode ke  $t$ . Model ARIMAX dengan efek variasi kalender dengan tren deterministik ditulis sebagai berikut:

$$\Phi_p(B)(1-B)^d Y_t = \beta_0 + \gamma_t + \beta_1 V_{1,t} + \dots + \beta_i V_{i,t} + \theta_q(B) \varepsilon_t \quad (2.19)$$

$Y_t$  : Data Pada Periode ke- $t$

$\gamma_t$  : Komponen deterministic ke- $t$

$\beta_i V_{i,t}$  : peubah boneka untuk efek variasi kalender ke- $p$

Lee, et.al [4], juga menuliskan prosedur untuk membangun model yang berpola variasi kalender sebagai berikut :

**langkah 1** : Menentukan variabel *dummy* untuk variasi kalender.

**langkah 2** : Memodelkan  $Z_t$  menggunakan regresi *timeseries* tanpa mengikutsertakan komponen musiman dan *trend* untuk membangun model ARIMAX dengan *stochastic trend*, dan atau memodelkan  $Z_t$  dengan mengikutkan keseluruhan komponen untuk membangun model ARIMAX dengan *deterministic trend*, untuk mendapatkan *error*  $a_t$

**langkah 3** : Memodelkan  $a_t$  menggunakan ARIMA.

**langkah 4** : Order ARIMA pada langkah 3, variabel *dummy* pada langkah 2 digunakan sebagai input secara simultan untuk persamaan (4) dan (6).

**langkah 5** : Melakukan uji signifikansi parameter, dan pengecekan asumsi *error* hingga asumsi *error* terpenuhi

## 2.6 Time Series Regression (TSR)

Regresi time series terdiri dari variabel dependen yang dipengaruhi oleh variabel independen yang telah diketahui. Berikut persamaan umum regresi menurut Draper & Smith [3].

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2,t} + \dots + \beta_k X_{k,t} + a_t \quad (2.20)$$

Koefisien parameter regresi pada persamaan di atas diperoleh dari perhitungan sebagai berikut

$$\beta = (X'X)^{-1}(X'Y) \quad (2.21)$$

dengan

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} \quad (2.22)$$

Model regresi time series dari data yang memiliki tren mengikuti persamaan berikut :

$$Y_t = \beta_1 t + \beta_2 D_{1,t} + \dots + \beta_j D_{it} + \beta_{j+1} t D_{1,t} + \dots + \beta_{j+k} t D_{it} + a_t \quad (2.23)$$

Dengan  $a_t$  adalah error yang telah memenuhi asumsi *white noise*. Pada model dengan pola musiman terdapat beberapa variabel dummy dengan asumsi bahwa ada M musiman (bulanan) per tahun dengan dummy musimannya  $(M_{1,t}, M_{2,t}, \dots, M_{12,t})$  yang dapat dituliskan sebagai berikut.

$$Y_t = \alpha_1 M_{1,t} + \alpha_2 M_{2,t} + \alpha_3 M_{3,t} + \dots + \alpha_{12} M_{12,t} + a_t \quad (2.24)$$

Dengan dummy  $M_{1,t}, M_{2,t}, \dots, M_{12,t}$  yang menyatakan pola musiman dengan periode dua belas bulan maka terdapat dua belas variabel dummy, satu variabel untuk satu bulan. Periode waktu berupa variasi kalender juga dapat dilakukan permodelan dengan menggunakan regresi. Model variasi kalender merupakan model time series yang digunakan untuk meramalkan data berdasarkan pola musiman dengan periode bervariasi. Model regresi untuk data dengan variasi kalender mengikuti persamaan sebagai berikut.

$$Y_t = \hat{\delta}_1 L_{1,t} + \hat{\delta}_2 L_{2,t} + \hat{\delta}_3 L_{3,t} + \hat{\delta}_4 L_{4,t} + a_t \quad (2.25)$$

Dan

$$Y_t = \hat{\gamma}_1 L_{1,t-1} + \hat{\gamma}_2 L_{2,t-1} + \hat{\gamma}_3 L_{3,t-1} + \hat{\gamma}_4 L_{4,t-1} + a_t \quad (2.26)$$

Dengan variabel dummy adalah  $L_{1,t}, L_{2,t}, L_{3,t}$ , dan  $L_{4,t}$  adalah efek hari raya yang terjadi pada minggu ke-1 sampai minggu ke-4. Untuk variabel dummy  $(L_{1,t-1}, L_{2,t-1}, L_{3,t-1}$  dan  $L_{4,t-1})$  adalah efek persiapan sebulan sebelum hari raya. Sementara  $a_t$  adalah error yang telah memenuhi asumsi *white noise*.

## 2.7 Evaluasi Model

Dalam penelitian ini, penentuan model terbaik menggunakan MAPE. MAPE merupakan gambaran ukuran ketepatan estimasi model maupun ramalan. Nilai MAPE pada data *training* digunakan untuk melihat ketepatan estimasi model sedangkan nilai MAPE pada data *testing* menggambarkan ketepatan ramalan model. MAPE dihitung dengan membagi persentase dari rata-rata harga mutlak residual pada tiap periode dengan nilai aktual. Menurut [8], MAPE dirumuskan sebagai :

$$\begin{aligned} MAPE &= \frac{1}{n} \sum_{t=1}^n \left| \frac{Z_t - \hat{Z}_t}{Z_t} \right| \\ &* 100\% \end{aligned} \quad (2.27)$$

dengan  $Z_t$  adalah nilai observasi pada waktu ke- $t$ ,  $\hat{Z}_t$  adalah nilai peramalan pada waktu ke- $t$  dan  $n$  adalah banyaknya observasi. Model yang baik memiliki nilai MAPE sesuai kriteria pada Tabel [5].

**Tabel 2.1.** Kriteria MAPE

Nilai MAPE (%)	Kriteria
< 10	Kemampuan peramalan sangat baik
10-20	Kemampuan peramalan baik
20-50	Kemampuan peramalan cukup
>50	Kemampuan peramalan buruk

### III. HASIL DAN PEMBAHASAN

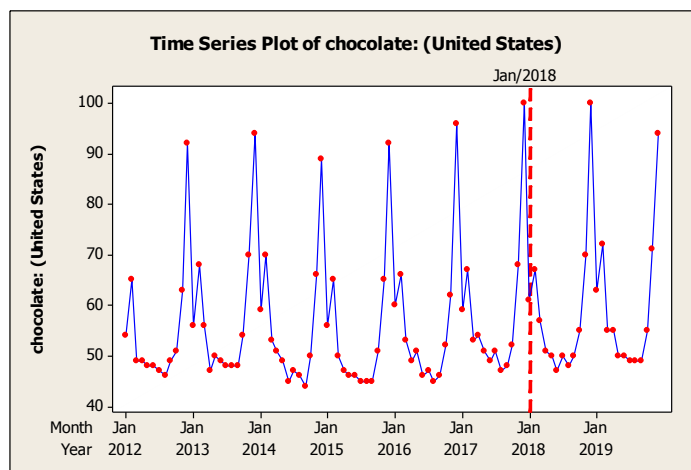
#### 3.1 Identifikasi Data

Data yang digunakan dalam penelitian ini diambil dari *google trend* yaitu dengan memasukkan kata kunci “Chocolate” untuk negara Amerika Serikat, dan “Cokelat” untuk negara Indonesia. Periode data yang digunakan yaitu Januari 2012 sampai Desember 2019. Dalam analisis, data dibagi menjadi 2 yaitu data *Training* dan data *Testing*. Data *Training* akan digunakan untuk mendapatkan model yang sesuai dengan karakteristik data, dan data *Training* digunakan untuk *forecasting* dan melihat akurasi nya. Rincian data yang digunakan sebagai berikut :

**Tabel 3.1** Rincian Data Cokelat yang digunakan dalam penelitian

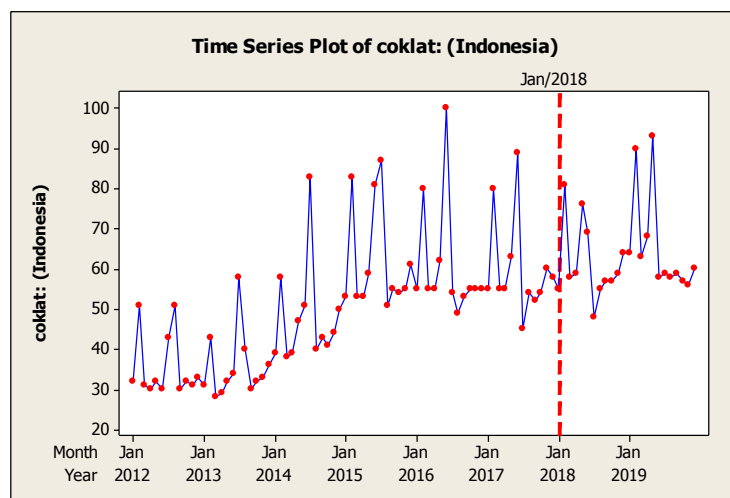
Negara	Total Data	Data Training	Data Testing
Indonesia	96	72	24
Amerika	96	72	24

Berdasarkan total data, dapat dilihat plot data cokelat di Indonesia dan Amerika Serikat per bulannya. Adapun *time series* plot nya sebagai berikut :



**Gambar 3.1.** *Time Series* Plot data Cokelat di Amerika Serikat





**Gambar 3.2.** *Time Series* Plot data Cokelat di Indonesia dan Amerika Serikat

Berdasarkan *Time Series* plot pada gambar 3.1 dan 3.2, pola data untuk data cokelat di Indonesia dan Amerika sama sama memiliki pola seasonal dan dipengaruhi oleh efek kalender variasi. Untuk Amerika Serikat, plot data memperlihatkan data yang tinggi pada bulan february bertepatan dengan tradisi *Valentine* dan pada bulan Desember bertepatan dengan hari Natal. Sedangkan untuk Indonesia, terlihat data cokelat menunjukkan data yang tinggi pada bulan february dan bulan hari raya Idul Fitri. Sehingga akan dilakukan proses analisis lanjutan. Analisis lanjutan akan menggunakan beberapa metode peramalan dan akan dilakukan perbandingan nilai MAPE dari setiap metode untuk memperoleh metode dengan nilai MAPE terkecil. Metode tersebut yaitu *Naïve Trend Linear*, *Naïve Trend Exponential*, *Double Exponential Smoothing*, *Time Series Regression*, ARIMA, dan ARIMAX. Untuk metode *Naïve Trend Linear*, *Naïve Trend Exponential*, dan *Double Exponential Smoothing* tidak dibahas secara rinci tetapi hanya dimunculkan sebagai pembanding terhadap metode lainnya.

### 3.2 *Time Series Regression*

Dalam melakukan peramalan menggunakan *Time Series Regression*, terlebih dahulu dibentuk variabel variabel *Dummy* untuk membantu pembentukan model *Time Series Regression*nya. Adapun variabel *Dummy* dalam penelitian ini sebagai berikut:

- Indonesia
  - $t$ ,  $D$ ,  $tD$  sebagai variabel *Dummy* untuk menangkap pola trend yang terjadi
  - $M1$ ,  $M2$ ,  $M3$ , ...,  $M12$  sebagai variabel *Dummy* yang menunjukkan bulan Januari sampai bulan Desember untuk menangkap efek musiman
  - $W1$ ,  $W2$ ,  $W3$ ,  $W4$  sebagai variabel *Dummy* yang menunjukkan minggu pertama sampai minggu keempat terjadinya Hari Raya Idul Fitri
  - $WB1$ ,  $WB2$ ,  $WB3$ ,  $WB4$  sebagai variabel *Dummy* yang menunjukkan minggu pertama sampai minggu keempat sebulan sebelum Idul Fitri untuk menangkap efek kalender variasi
- Amerika Serikat
  - $t$ ,  $D$  sebagai variabel *Dummy* untuk menangkap pola trend yang terjadi
  - $M1$ ,  $M2$ ,  $M3$ , ...,  $M12$  sebagai variabel *Dummy* yang menunjukkan bulan Januari sampai bulan Desember untuk menangkap efek musiman

Sehingga dapat dilakukan pembentukan model *Time Series Regression* dari masing masing data cokelat di Indonesia dan Amerika Serikat.

### 3.3 Pemodelan ARIMA

Langkah-langkah pemodelan ARIMA untuk data coklat di Indonesia dan di Amerika Serikat sebagai berikut :

- Melakukan identifikasi terhadap data coklat Indonesia dan Amerika Serikat.
- Melakukan pembagian data *Training* dan *Data Testing*
- Pengecekan stasioner terhadap varians dari masing masing data menggunakan *box cox plot*. Jika data belum stasioner, dilakukan transformasi terhadap data
- Pengecekan stasioner terhadap rata-rata dengan melihat plot ACF dan PACF dari masing masing data coklat di kedua negara.
- Pembentukan model ARIMA
- Melakukan uji signifikansi parameter
- Melakukan pengujian asumsi *white noise*, dan kenormalan terhadap residual
- Penentuan model terbaik dengan melihat nilai MAPE atau RMSE terkecil dari residual nya

### 3.4 Pemodelan ARIMAX

Peramalan terhadap data yang memiliki trend dan dipengaruhi oleh kalender variasi harus dilakukan dengan menggunakan metode gabungan atau *Hybrid*. Dalam kasus data coklat di Indonesia dan di Amerika Serikat, akan diterapkan metode ARIMAX dengan efek kalender variasi untuk melakukan peramalan. Hasil peramalan akan dibandingkan dengan metode klasik lainnya dan juga metode ARIMA. Adapun langkah-langkah pemodelan ARIMAX yaitu :

- Menentukan variabel *dummy* untuk periode variasi kalender
- Menguji kestasioneran data *training*
- Mengeluarkan pengaruh variasi kalender dengan memodelkan data *insample* yang telah stasioner dengan variabel *dummy* untuk memperoleh residual
- Melakukan pemodelan terhadap residual yang diperoleh dengan metode ARIMA.
- Membangun model ARIMAX dari data *training* dengan menggunakan orde dari model ARIMA pada langkah sebelumnya.
- Melakukan uji signifikansi parameter.
- Melakukan pengujian asumsi diagnostik pada residual dengan menggunakan uji *White Noise* dan uji normalitas
- Melakukan pemilihan model terbaik dengan melihat nilai MAPE atau RMSE residual.

### 3.5 Perbandingan Akurasi dan Peramalan

Untuk mengetahui metode terbaik yang dapat digunakan untuk melakukan peramalan, maka akan dilakukan perbandingan nilai MAPE untuk setiap metode yang digunakan untuk data coklat di kedua negara. Metode metode yang digunakan dalam peramalan data coklat ini yaitu metode *Naïve Trend Linear*, *Naïve Trend Exponensial*, *Double Exponential Smoothing*, *Time Series Regression*, *ARIMA*, dan *ARIMAX With Calender Effect*. Ringkasan perbandingan nilai MAPE dari semua metode yang digunakan dapat dilihat pada tabel berikut :

**Tabel 3.2** Nilai MAPE setiap metode terhadap data Cokelat di Indonesia

Metode	Training	Hasil Forecast Indonesia				
		1	3	6	12	24
<i>Naïve Trend Linear</i>	40.98	0.018	15.165	21.821	25.993	46.977

<i>Naïve Trend Exponensial</i>	41.24	0.019	14.899	20.976	22.884	37.194
<i>Double Exponential Smoothing</i>	21.86	7.438	12.719	11.487	14.278	18.621
<i>Time Series Regression</i>	5.40	<b>0.546</b>	4.922	<b>8.393</b>	8.628	10.593
ARIMA (0,1,[1,13])(1,1[2]) <sup>12</sup>	5.49	11.007	5.835	7.954	13.320	17.197
ARIMAX (1,0,0)	<b>4.73</b>	1.024	<b>4.311</b>	8.438	<b>8.081</b>	<b>9.299</b>

**Tabel 3.3** Nilai MAPE setiap metode terhadap data Cokelat di Amerika Serikat

Metode	Training	Hasil Forecast Indonesia				
		1	3	6	12	24
<i>Naïve Trend Linear</i>	24.06	116.3 9	168.34	298.89	437.69	772.21
<i>Naïve Trend Seasonal</i>	4.48	3.27	3.43	3.74	3.24	3.66
<i>Double Exponential Smoothing</i>	14.94	56.81	57.49	79.96	79.44	88.98
<i>Time Series Regression</i>	2.94	4.25	3.97	<b>2.64</b>	3.22	3.34
ARIMA (2,1,0)(0,1,0) <sup>12</sup>	3.41	<b>1.61</b>	3.08	7.56	6.15	8.96
ARIMAX ([3],0,0)	<b>2.47</b>	3.91	<b>2.47</b>	2.87	<b>2.69</b>	<b>2.64</b>

Berdasarkan data nilai MAPE yang disajikan pada tabel 3.2 dan tabel 3.3, metode metode yang digunakan dalam penelitian ini menghasilkan nilai MAPE yang beragam. Untuk Indonesia, metode terbaik berdasarkan data *Training*, metode ARIMAX (1,0,0) with Calendar Variation Effect memiliki nilai MAPE yang paling kecil yaitu 4.37. Nilai MAPE tersebut menunjukkan bahwa metode tersebut memiliki kemampuan yang sangat baik dalam melakukan peramalan.

Berdasarkan nilai MAPE hasil peramalan 1 bulan kedepan, 3 bulan kedepan, 6 bulan kedepan, 12 bulan kedepan dan 24 bulan kedepan, terdapat 2 metode yang memiliki kemampuan peramalan yang sangat baik, yaitu metode *Time Series Regression* untuk peramalan 1 bulan kedepan dan 6 bulan kedepan serta metode ARIMAX (1,0,0) with Calendar Variation Effect untuk peramalan 3 bulan kedepan, 12 bulan kedepan dan 24 bulan kedepan. Hal ini menunjukkan bahwa model terbaik pada data *Training* tidak selalu menghasilkan nilai MAPE yang terbaik juga untuk data *Forecast* nya.

Sedangkan data cokelat di Amerika Serikat yang disajikan pada tabel 3.3 menunjukkan bahwa metode terbaik yang terbentuk menggunakan data *Training* adalah metode ARIMAX ([3],0,0) dengan nilai MAPE mencapai 2.47. Hal ini menyimpulkan bahwa metode tersebut memiliki kemampuan yang sangat baik dalam melakukan peramalan.

Untuk data hasil forecast-nya 1 bulan sampai 24 bulan kedepan, terdapat 3 metode yang memberikan nilai MAPE terkecil yaitu *Time Series Regression*, ARIMA (2,1,0)(0,1,0)<sup>12</sup>, dan ARIMAX ([3],0,0). *Time Series Regression* baik digunakan untuk peramalan 6 bulan kedepan, ARIMA (2,1,0)(0,1,0)<sup>12</sup> baik digunakan untuk peramalan 1 bulan kedepan, sedangkan ARIMAX ([3],0,0) baik digunakan untuk peramalan 3 bulan kedepan, 12 bulan kedepan dan 24 bulan kedepan. Hal ini juga menunjukkan bahwa secara umum, metode terbaik yang diperoleh dari data *Training* tidak selalu memberikan nilai MAPE terbaik juga untuk data *testing*-nya.

Data cokelat yang diperoleh melalui *Google Trend* untuk kasus di Indonesia dan Amerika dapat memberikan gambaran mengenai peramalan suatu data yang memiliki pola

trend dan terdapat efek kalender variasi yang membuat data memiliki pola yang berbeda setiap tahunnya.

#### 4. KESIMPULAN

Berdasarkan uraian pada pembahasan sebelumnya, dapat disimpulkan beberapa poin sebagai berikut :

- Data pada *google trend* dapat menjadi alternatif dalam penyediaan data yang bersifat runtun waktu. data *Google Trend* mencerminkan penelusuran yang dilakukan pengguna di Google terhadap topik tertentu dan dapat digunakan untuk melakukan suatu kajian atau penelitian.
- Model peramalan data cokelat di Indonesia dan di Amerika Serikat berbeda. Data cokelat di Amerika Serikat memiliki pola trend dan meningkat pada bulan Desember dan Februari setiap tahunnya. Bulan tersebut bertepatan dengan Natal pada bulan Desember dan Tradisi *Valentine* pada bulan Februari sehingga penelusuran data cokelat meningkat.
- Di Indonesia, penelusuran data cokelat memiliki pola trend dan meningkat pada bulan Februari dan bulan terdapatnya hari raya Idul Fitri. Sehingga diperlukan pemodelan yang bersifat Efek Kalender Variasi untuk peramalannya.
- Peramalan terhadap data penelusuran Cokelat di Indonesia dan Amerika Serikat menggunakan metode *ARIMAX with Calendar Variation Effect*.
- Metode *ARIMAX with Calendar Variation Effect* menghasilkan nilai MAPE yang sangat presisi dalam melakukan peramalan terhadap data cokelat di Indonesia dan Amerika Serikat. Berdasarkan hasil penelitian, nilai MAPE yang dihasilkan berada di bawah 10 persen, sehingga dapat disimpulkan metode tersebut memiliki kemampuan yang sangat baik dalam melakukan peramalan.
- Jika dibandingkan dengan metode lain seperti metode *Naïve Trend Linear*, *Naïve Trend Exponensial*, *Double Exponential Smoothing*, *Time Series Regression*, dan *ARIMA*, metode *ARIMAX with Calendar Variation Effect* memberikan nilai MAPE yang lebih baik pada data *Training* di Indonesia dan Amerika Serikat.
- Berdasarkan hasil ramalannya kedepan untuk Indonesia, metode *Time Series Regression* sangat baik digunakan untuk peramalan 1 bulan kedepan dan 6 bulan kedepan sedangkan metode *ARIMAX (1,0,0) with Calendar variation Effect* untuk peramalan 3 bulan kedepan, 12 bulan kedepan dan 24 bulan kedepan.
- Berdasarkan hasil ramalannya kedepan untuk Amerika Serikat, metode *Time Series Regression* baik digunakan untuk peramalan 6 bulan kedepan, *ARIMA (2,1,0)(0,1,0)*<sup>12</sup> baik digunakan untuk peramalan 1 bulan kedepan, sedangkan *ARIMAX ([3],0,0)* baik digunakan untuk peramalan 3 bulan kedepan, 12 bulan kedepan dan 24 bulan kedepan.
- Nilai MAPE yang presisi dari suatu metode pada data *Training* tidak selalu memberikan jaminan akan menghasilkan nilai MAPE yang presisi juga pada data ramalan yang dihasilkan.

#### DAFTAR PUSTAKA

- [1] Anggraeni, W., Retno A.V., Yuni D.K., 2015. Performance Comparisons Between Arima and Arimax Method in Moslem Kids Clothes Demand Forecasting: Case Study. *Procedia Computer Science* 72 (2015), 630 – 637.

- [2] Farih, N.L., Irma F., Madona W., 2019. Prediction of the Number of Ship Passengers in the Port of Makassar Using ARIMAX method in the Presence of Calendar Variation. *Indonesian Journal of Pure and Applied Mathematics*, Vol. 1, No. 1 (2019), pp. 57-67.
- [3] Hanim, Yulia Masnunah, 2015. Penerapan Regresi Time Series dan ARIMAX untuk Peramalan Inflow dan Outflow Uang Kartal di Jawa Timur, DKI Jakarta dan Nasional. *Institut Teknologi Sepuluh Nopember*.
- [4] Lee, M. H., Suhartono, dan Hamzah, N.A., 2010, Calendar variation model based on ARIMAX for forecasting sales data with Ramadhan effect. *Proc. RCSS'10, Malaysia*, 349-361.
- [5] Purnama, D.I 2020. Model Hybrid Seasonal Autoregressive Integrated Moving Average (Sarima)-Support Vector Regression (Svr) Dan Penerapannya Pada Harga Gabah Nasional. *Tesis : Universitas Padjadjaran, Fakultas Matematika dan Ilmu Pengetahuan Alam / Statistika Terapan*.
- [6] S.Makridakis, S. C.Wheelwright dan V. E. McGee, 1999. Metode dan Aplikasi Peramalan Jilid 1 Edisi Revisi (terj.), *Alih Bahasa: Hari Suminto*, Jakarta: Binapura Aksara.
- [7] Spillane, J. J. 1995. Komoditi kakao: peranannya dalam perekonomian Indonesia. Cet. 1. Yogyakarta: Kanisius.
- [8] Wei, W.W.S. 2006. Time Series Analysis: Univariate and Multivariate Methods, Second Edition. *New York : Pearson Education*.
- [9] Utama, Made Suryana, and I. Gusti Putu Nata Wirawan, 2014. Model Box-Jenkins dalam Rangka Peramalan Produk Domestik Regional Bruto Provinsi Bali. *Buletin Studi Ekonomi*.
- [10] Wulansari, E. R., dan Suhartono, 2014. Peramalan Netflow Uang Kartal Dengan Metode ARIMAX dan Radial Basis Function Network (Studi Kasus Di Bank Indonesia). *Jurnal Sains dan Seni POMITS*. No.2, pp. 73-78.
- [11] Yударuddin, Rizky, 2019. Forecasting untuk Kegiatan Ekonomi dan Bisnis. *RV Pustaka Horizon*.