



IPB University
Bogor Indonesia

Data Sciences

MAKALAH TUGAS (2021)

Mata Kuliah MAT523 –Komputasi Matematika

IPB University

Matematika Terapan
Sekolah Pascasarjana

*"Ilmu itu kekayaan meski tanpa harta,
kemuliaan meski tanpa pendukung,
dan kekuasaan meski tanpa
pasukan."*

(Abu Ja'far Ath-Thahawi)

Daftar Isi

Bab 1. Data Saham

Analisis Model ARCH/GARCH pada Pemodelan Harga Saham Bulanan PT. Wijaya Karya (Persero) Tbk.

periode 2010-2020 oleh Dillah Rismawati	1
Analisis Harga Saham WIKA Tahun 2010-2020 dengan Metode <i>Simple Moving Average</i> dan <i>Exponential Moving Average</i> oleh Fitra Nuvus Salsabila	11
Analisis Data Saham Menggunakan Metode <i>K-Means Clustering</i> dengan Visualisasi Data dalam Bentuk <i>Plot/Scatter</i> dan <i>Boxplot</i> oleh Yohanes Mario Defianus Beti	21
Pengaplikasian Metode <i>Autoregressive</i> Dalam Memprediksi Data Saham oleh Ade Irawan	29

Bab 2. Data Curah Hujan Bulanan di Kalimantan

Estimasi <i>Trend and Seasonal Component</i> Data Curah Hujan Beberapa Kota di Kalimantan oleh Refi Revina	35
Perbandingan Hasil <i>Clustering</i> Curah Hujan di Wilayah Kalimantan dengan Metode <i>K-Means</i> dan <i>K-Medoids</i> oleh Adelia Christi	45
Analisis Curah Hujan Rencana Maksimum Kota Sendawar, Banjarmasin, dan Kotabaru Menggunakan Metode Gumbel oleh Evi Ardiyani	51
Analisis Curah Hujan di Pulau Kalimantan Menggunakan Metode PCA (<i>Principal Component Analysis</i>) oleh Salsabilla Rahmah	59

Bab 3. Data Inflasi

Penerapan Metode <i>K-Means</i> dan <i>K-Medoids Clustering</i> pada Laju Inflasi Beberapa Kategori di Indonesia oleh Siti Nurhayatudin	65
Meramalkan Perubahan Inflasi Pada Tahun 2020 Setiap Bulannya oleh Rahmawati Masithoh	71
Analisis ARIMA untuk Menentukan Model Peramalan Tingkat Inflasi oleh Anggun Praptaningsih	77

Bab 4. Data Cuaca Harian di Jakarta

Analisis Faktor yang Mempengaruhi Cuaca di Jakarta Tahun 2019 dengan Metode <i>Principal Component Analysis</i> (PCA) oleh Yusti Qomah	85
Klasterisasi Curah Hujan Menggunakan Metode <i>K-Means</i> dan <i>K-Medoids</i> (Studi Kasus di Jakarta Tahun 2019) oleh Euis Aprianti	93
Analisis Pengaruh Suhu Udara Terhadap Kelembaban Udara di Jakarta pada Tahun 2019 oleh Trianty Putri Blante	99

Prakata

Buku elektronik (*e-book*) ini merupakan kumpulan makalah dari tugas *mini-project* mahasiswa Mata Kuliah MAT523 Komputasi Matematika semester ganjil tahun akademik 2021/2022. Tugas *mini-project* bertujuan untuk meningkatkan pemahaman, kreativitas, dan kemampuan penyelesaian masalah mahasiswa dalam menerapkan metode-metode *data sciences* yang telah dipelajari, seperti pembacaan dan pendeskripsian data, aljabar linear, statistik, reduksi dimensi, *clustering*, klasifikasi, regresi, visualisasi, dan metode terkait lainnya. Mahasiswa diberikan empat jenis data dari sumber yang berbeda dan setiap mahasiswa memperoleh satu dari empat data tersebut untuk diolah dan dipresentasikan. Keempat jenis data tersebut adalah

- Data harga dan imbal hasil dari empat saham di Indonesia pada tahun 2010-2020
- Data curah hujan bulanan di 15 kota berbeda di Kalimantan pada tahun 2001-2019
- Data inflasi di Indonesia berdasarkan beberapa kategori pada tahun 1990-2019
- Data cuaca (suhu, kelembapan, kecepatan dan arah angin) per jam di Jakarta pada tahun 2019

Mini-project berlangsung selama empat minggu mulai dari tanggal 24 September – 24 Oktober 2021. Luaran dari tugas *mini-project* ini berupa *file notebook interactive* (ipynb) yang dipresentasikan pada tanggal 18 Oktober 2021 dan berupa makalah yang dikumpulkan pada tanggal 24 Oktober 2021. *File notebook interactive* (ipynb) yang berisi pemrograman untuk menyelesaikan masalah dapat diakses pada laman <https://github.com/mkhoirun-najiboi/kom-mat-2021>.

Semoga buku elektronik ini dapat memberikan manfaat dan meningkatkan pemahaman penerapan metode *data sciences* lebih dalam. Kritik dan saran sangat diharapkan untuk perkembangan praktikum di masa yang akan datang. Terima kasih.

Bogor, 25 Oktober 2021

Asisten Praktikum

Mohamad Khoirun Najib, S.Si

NIM. G5501201007

Analisis Model ARCH/GARCH pada Pemodelan Harga Saham Bulanan Wijaya Karya (Persero) Tbk. periode 2010-2020

Dillah Rismawati

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: dillah_rismawati@apps.ipb.ac.id

Abstrak.

Peramalan harga saham sangat dibutuhkan bagi para investor sebagai informasi fluktuasi harga saham di masa yang akan datang. Salah satu metode yang sering digunakan dalam peramalan (*forecasting*) harga saham adalah *Autoregressive Integrated Moving Average* (ARIMA). Harga saham yang berfluktuasi seiring dengan berjalanannya waktu, akan menyebabkan terjadinya heteroskedastisitas (ragam tidak homogen). Metode yang dapat digunakan untuk mengatasi ragam yang tidak konstan tersebut adalah model *Autoregressive Conditional Heteroskedasticity* (ARCH) dan *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH). Data yang digunakan dalam projek ini adalah data harga saham bulanan Wijaya Karya (Persero) Tbk periode 2010-2020. Tujuan projek ini adalah menganalisis model ARCH/GARCH pada harga saham PT. Wijaya Karya Tbk. dan peramalannya untuk 12 periode. Hasil evaluasi dan validasi model menunjukkan model yang dapat digunakan untuk data harga saham PT. WIKA Tbk. adalah model ARIMA(2,1,2)-GARCH(1,1). Hasil forecasting selama 12 periode ke depan menunjukkan hasil yang kurang fit dengan data asli. Hal ini dapat disebabkan karena model yang dipakai belum memenuhi kriteria model yang terbaik.

Kata Kunci: ARCH-GARCH , ARIMA, peramalan

1. Pendahuluan

Saham adalah dokumen berharga yang menunjukkan bagian kepemilikan atas suatu perusahaan. Dengan kata lain, ketika seseorang membeli saham maka orang tersebut telah membeli sebagian kepemilikan atas perusahaan tersebut (Prawiro 2018). Harga saham suatu perusahaan dapat berpengaruh pada harga saham pasar, karena harga pasar terbentuk dari gabungan harga saham individu yang memperdagangkan sahamnya di pasar modal (Puspitaningtyas 2015). Harga saham individu dan harga pasar digunakan untuk memprediksi return saham individu dan return pasar sebagai informasi untuk memprediksi koefisien β (risiko sistematis). Keuntungan dari mengetahui resiko adalah dapat mengubah perilaku pemilik saham untuk meminimumkan terjadinya risiko. Resiko kerugian yang diperoleh investor adalah *capital loss* yang merupakan kebalikan dari *capital gain*. Capital loss terjadi ketika investor menjual harga saham lebih rendah daripada harga beli. Selain mempertimbangkan faktor resiko, investor dalam berinvestasi di pasar saham juga mengumpulkan informasi sebanyak mungkin, termasuk informasi yang berkaitan dengan harga saham dan kinerja perusahaan (Wahyudi, 2017). Oleh karena itu, dibutuhkan suatu pemahaman mengenai harga saham itu sendiri untuk saat ini atau dalam jangka panjang yang salah satu diantaranya dapat diamati melalui pemodelan harganya.

Pemodelan harga saham dapat dilakukan dengan cara melakukan peramalan (*forecasting*) untuk waktu yang akan datang. Salah satu metode yang sering digunakan dalam peramalan adalah *Autoregressive Integrated Moving Average* (ARIMA). Model ARIMA terkenal karena keakurasiannya namun memiliki keterbatasan utama yaitu hanya dapat meramalkan model prediksi linier dan tidak sesuai untuk model prediksi nonlinier yang kompleks (Zhang & Qi, 2005). Selain itu, keterbatasan yang lain adalah model ARIMA tidak mampu mengatasi variansi error pada data time series finansial yang mempunyai tingkat volatilitas yang cukup tinggi (Soeksir dan Fatanah 2020). Harga saham yang berfluktuasi baik terus meningkat atau menurun dengan seiring berjalanannya waktu, akan menyebabkan ragamnya terus meningkat pula. Kondisi tersebut ada kemungkinan dapat menyebabkan terjadinya heteroskedastisitas atau ragam tidak homogen. Heteroskedastisitas menyebabkan pemodelan dan peramalan dengan metode peramalan standar tidak dapat lagi diaplikasikan pada data saham (Khoirunnisa 2014). Untuk mengatasi hal tersebut, salah satu model yang dapat digunakan adalah model *Autoregressive Conditional Heteroskedasticity* (ARCH) dan *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH) untuk mengatasi variansi error yang tidak konstan (keragaman tinggi).

Tujuan dari project ini adalah menganalisis model ARCH/GARCH pada harga saham PT. Wijaya Karya Tbk. Analisis ini akan diawali dengan pembentukan model ARIMA. Ratnasari (2014), menjelaskan bahwa model ARCH/GARCH pada metode ARIMA cocok digunakan dalam pemodelan pada data harga saham yang memiliki resiko dan fluktuasi tinggi. Setelah didapatkan model terbaik akan dilakukan *forecasting* harga saham tersebut untuk 12 periode ke depan.

2. Data dan Metode

2.1. Alat dan Bahan

Alat dan bahan yang digunakan dalam makalah ini adalah *personal computer* (PC) dengan dilengkapi processor intel Pentium silver N5030, memory yang digunakan 8 GB DDR4, dan penyimpanan 512 GB. System operasi yang digunakan adalah Windows 10. Selain itu, makalah ini juga dibantu dengan menggunakan *software Julia* versi 1.6.2. Adapun paket-paket yang digunakan dalam project ini tertera pada Tabel 1.

Tabel 1. Paket-paket di Julia yang digunakan dalam Project

Nama Paket	Fungsi
<i>Plots.jl</i>	Memvisualisasikan data yang digunakan dalam bentuk plot
<i>CSV.jl</i>	Membaca file berformat csv dalam Julia
<i>DataFrames.jl</i>	Menyediakan seperangkat alat untuk bekerja dengan data tabular di Julia
<i>StatsBase.jl</i>	Memanggil fungsi <i>describe()</i> digunakan untuk menjelaskan spesifikasi masing-masing data
<i>Statistics.jl</i>	Menggambarkan beberapa gambaran statistik data
<i>VegaDataset.jl</i>	Menyediakan dataset yang tersedia di vega-dataset dari Julia
<i>Vegalite.jl</i>	Membuat berbagai plot statistic yang mengakses fungsi dari Vega-Lite yang mendasarinya dan terintegrasi erat ke dalam ekosistem Julia
<i>HypothesisTest.jl</i>	Mengimplementasikan beberapa uji hipotesis di Julia
<i>ARCHModels.jl</i>	Mengimplementasikan simulasi, estimasi, <i>forecasting</i> , dan <i>testing</i> untuk berbagai model
<i>Turing.jl</i>	Mengintegrasikan model menggunakan sintax Julia standar, dan menyediakan berbagai metode inferensi berbasis <i>sampling</i> untuk memecahkan masalah di <i>machine learning</i> , statistik Bayesian, dan <i>data science</i> yang berfokus fokus khusus pada modularitas, dan memisahkan bahasa pemodelan (yaitu <i>compiler</i>) dan metode inferensi
<i>TSAnalysis.jl</i>	Menyediakan <i>tools</i> dasar yang digunakan untuk analisis data <i>time series</i> seperti ARIMA, VARIMA, dan <i>Kalman filter and smoother</i>
<i>Optim.jl</i>	Optimisasi berbagai fungsi
<i>Measures.jl</i>	Membangun dan mewakili ekspresi yang melibatkan berbagai jenis unit yang kemudian dievaluasi, menyelesaikannya menjadi unit absolut.

2.2. Data

Data yang digunakan dalam makalah ini adalah data Harga saham bulanan Wijaya Karya (Persero) Tbk periode 2010-2020 Wijaya Karya (Persero) Tbk., yang bersumber dari <https://finance.yahoo.com>. Data yang digunakan untuk pendugaan model adalah data periode Januari 2010-Desember 2019, yaitu sebanyak 120 data. Sedangkan 12 bulan data terakhir akan digunakan untuk validasi model hasil *forecasting* dengan ARIMA.

2.3. Metode

Metode yang digunakan dalam project ini adalah metode *Autoregressive Integrated Moving Average* (ARIMA) dengan model ARCH/GARCH yang sesuai.

- Model *autoresgressive* orde ke-*p* AR(*p*)

Model *Autoregressive* (AR) merupakan model yang dipengaruhi oleh nilai-nilai masa lalu dari runtun waktu. Bentuk umum model AR(*p*) diberikan oleh :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + e_t \quad (1)$$

Keterangan :

ϕ_i = Koefisien AR ke-*i*

e_t = error kesalahan saat t, $e_t \sim WN(0, \sigma^2)$

(Ruey 2005).

- Model *Moving Average* orde ke-*q* MA(*q*)

Model *Moving Average* (MA) merupakan model yang dipengaruhi oleh residual masa lalu dari runtun waktu. Bentuk umum model MA(*q*) :

$$X_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (2)$$

dimana :

e_t = residual waktu ke -t

θ_i = koefisien MA orde *q*, $e_t \sim WN(0, \sigma^2)$

(Ruey 2005).

3. Model *autoresgressive Moving Average ARMA(p,q)*

Model *autoresgressive Moving Average* (ARMA) adalah model campuran proses AR dan MA. Bentuk umum model ARMA(p,q) :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (3)$$

Keterangan :

ϕ_i = Koefisien AR ke-*i*

θ_i = koefisien MA orde *q*

e_t = residual waktu ke *t*, $e_t \sim WN(0, \sigma^2)$

(Ruey 2005)

4. Model *Autoresgressive Integrated Moving Average ARIMA(p,d,q)*

Model ARIMA merupakan model campuran AR dan MA setelah dilakukan *differencing*. Bentuk umum model ARIMA adalah sebagai berikut.

$$(1 - B)^d X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (4)$$

(Ruey 2005)

5. Model *Autoregressive Conditional Heteroscedasticity* ARCH(p)

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \cdots + \alpha_i e_{t-j}^2 \quad (5)$$

(Tsay 2002)

6. Model *Generalized Autoregressive Conditional Heteroscedasticity* GARCH(p,q)

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i e_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (6)$$

(Tsay 2002)

2.4. Tahapan Pemodelan

Langkah-langkah yang dilakukan untuk menganalisis model ARCH/GARCH pada harga saham PT. Wijaya Karya Tbk. dan peramalannya untuk 12 periode adalah sebagai berikut.

1. Identifikasi model

Membuat plot data harga saham PT. WIKA, Tbk. untuk melihat kestasioneran data dalam rataan dan ragamnya.

2. Uji Akar Root menggunakan *Augmented Dickey-Fuller Test* (ADF Test)

Hipotesis :

$H_0: \gamma = 0$ (Terdapat unit root, sehingga data tidak stasioner dalam rataan)

$H_1: \gamma = 0$ (Tidak terdapat unit root, sehingga data stasioner dalam rataan)

Terdapat 2 kondisi :

- a) Jika data sudah stasioner, maka lanjut ke tahap selanjutnya
- b) Jika data belum stasioner, maka perlu dilakukan penstasioneran terlebih dahulu dengan melakukan pembedaan (*differencing*)

3. Identifikasi *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF)

Membuat plot ACF dan PACF untuk mengestimasi parameter AR dan MA.

Tabel 2. Identifikasi *p* dan *q* melalui nilai ACF dan PACF

	AR(<i>p</i>)	MA(<i>q</i>)	ARMA(<i>p, q</i>), <i>p > 0</i> dan <i>q > 0</i>
ACF	Tails off (menurun secara perlahan)	Cuts off after lag <i>q</i> (menurun secara drastis setelah lag <i>q</i>)	Tails off (menurun secara perlahan)
PACF	Cuts off after lag <i>p</i> (menurun secara drastis setelah lag <i>p</i>)	Tails off (menurun secara perlahan)	Tails off (menurun secara perlahan)

Sumber : Cryer and Chan (2008)

4. Uji Bartlett (Kehomogenan Ragam)

5. Pembentukan model ARIMA

Menentukan orde dari ARIMA untuk model pendahuluan. Selanjutnya Melakukan pendugaan parameter model ARIMA dengan melihat signifikansi parameternya. Pemilihan model ARIMA yang digunakan akan dilihat dari nilai *root mean square error* (RMSE) terkecil yaitu

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n}}$$

y_i = data actual

\bar{y} = data prediksi

n = jumlah data

6. Verifikasi model ARIMA

Asumsi yang harus dipenuhi adalah :

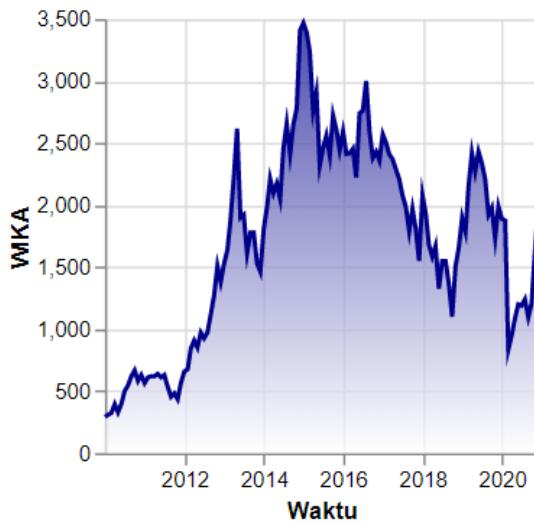
1. Kenormalan sisaan (uji *JarqueBera*)
2. Kehomogenan ragam (dilihat dari plot ACF dan PACF)
3. Kebebasan sisaan (uji *Ljung-Box*)
7. Peramalan (*forecasting*) data saham PT WIKA Tbk. untuk 12 bulan kedepan
8. Pendugaan parameter ARCH dan GARCH

Melihat berapa lag dari uji LM yang signifikan / mengandung pengaruh ARCH. Model ARCH spesifik untuk ordo rendah (Gujarati 1997), sehingga jika terdapat banyak lag yang signifikan yang menyebabkan model tidak efisien maka dibutuhkan perluasan dari model ARCH yaitu GARCH.

9. Menduga parameter dari GARCH dengan melihat keseluruhan parameternya signifikan. Parameter diartikan signifikan apabila nilai-p < taraf nyata 5%.
10. Menghitung volatilitas.

3. Hasil dan Pembahasan

Data pengamatan yang digunakan adalah data harga saham PT. Wijaya Karya Tbk. periode Januari 2010 hingga Desember 2019. Data periode Januari 2020 sampai Desember 2020 akan digunakan untuk validasi model hasil peramalan yang akan dilakukan dengan ARIMA. Plot data *time series* harga saham WIKA terdapat pada Gambar 1. Dari Gambar 1 terlihat bahwa data saham tersebut tidak stasioner yang ditunjukkan oleh fluktuasi harga yang tidak menyebar dalam rataan yang sama.



Gambar 1. Plot Harga Saham WIKA Periode 2010-2019

3.1 Uji Kestasioneran Data

Untuk memastikan bahwa Gambar 1 tidak stasioner, dapat dilakukan uji *Augmented Dickey-Fuller Test* (ADF Test) untuk mengecek kestasioneran data tersebut. Seddighi *et al* (2000) menjelaskan hipotesis dari pengujian ini adalah sebagai berikut.

$H_0: \gamma = 0$ (Terdapat unit root, sehingga data tidak stasioner dalam rataan)

$H_1: \gamma = 0$ (Tidak terdapat unit root, sehingga data stasioner dalam rataan)

H_0 ditolak apabila nilai statistik dari uji akar unit lebih besar daripada nilai kritis MacKinnon sehingga data disimpulkan data tidak stasioner. Dengan menggunakan paket HypothesisTest.jl didapatkan hasil sebagai berikut.

```

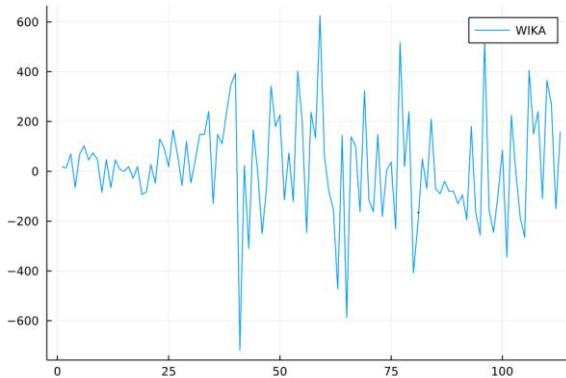
Augmented Dickey-Fuller unit root test
-----
Population details:
  parameter of interest: coefficient on lagged non-differenced variable
  value under h_0:          0
  point estimate:         -0.0465946

Test summary:
  outcome with 95% confidence: fail to reject h_0
  p-value:                   0.4022

Details:
  sample size in regression:      108
  number of lags:                 5
  ADF statistic:                  -1.75659
  Critical values at 1%, 5%, and 10%: [-3.48906 -2.88725 -2.58048]

```

Hasil uji ADF tersebut menunjukkan gagal untuk menolak H_0 atau terima H_0 , maka dapat disimpulkan bahwa data harga saham tersebut belum stasioner. Data tersebut perlu dilakukan penstasioneran terlebih dahulu dengan melakukan *differencing*. Proses difference atau perbedaan yang berordo-dilakukan jika data deret waktu bersifat nonstasioner, karena aspek-aspek AR dan MA dari model ARIMA menghendaki data harus bersifat stasioner. Sisaan dalam model didasarkan pada *white noise* atau galat bersifat acak (Montgomery et al.2008). Setelah dilakukan *differencing*, didapatkan plot baru sebagai berikut. Terlihat bahwa data sudah menyebar dalam rataan.



Gambar 2. Plot Harga saham setelah dilakukan *differencing* satu kali

Hasil uji ADF setelah dilakukan *differencing* satu kali adalah menolak hipotesis H_0 , artinya data sudah tidak terdapat *unit root* seingga data sudah stasioner.

```

Augmented Dickey-Fuller unit root test
-----
Population details:
  parameter of interest: coefficient on lagged non-differenced variable
  value under h_0:          0
  point estimate:         -1.08754

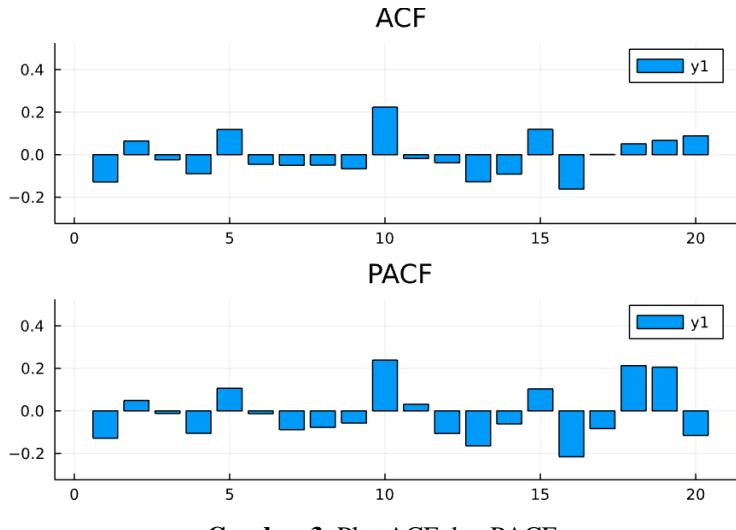
Test summary:
  outcome with 95% confidence: reject h_0
  p-value:                   0.0013

Details:
  sample size in regression:      107
  number of lags:                 5
  ADF statistic:                  -4.02529
  Critical values at 1%, 5%, and 10%: [-3.48959 -2.88748 -2.5806]

```

Setelah data menjadi stasioner, langkah selanjutnya adalah mengestimasi nilai parameter AR(p) dan MA(q) dari plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF). Plot ACF dan PACF digunakan untuk menentukan orde dari modelnya. ACF berfungsi untuk mengukur korelasi antar pengamatan dengan jeda k , sedangkan PACF mengukur korelasi dengan jeda k dan mengontrol korelasi antar dua pengamatan dengan jeda kurang dari k .

(Kelikume & Salami 2014). Sebelumnya telah diketahui bahwa data stasioner setelah dilakukan pembedaan (differencing) satu kali, sehingga dapat diketahui orde $d = 1$. Plot ACF dan PCF ditunjukkan oleh Gambar 3. Plot ACF dan PACF diatas menunjukkan bahwa nilai autokorelasi plot ACF data harga saham WIKA 120 periode terdapat cuts off setelah lag 1, dan pada plot PACF terlihat bahwa nilai autokorelasi parsial cuts off lag 1, sehingga nilai parameter Ma dan AR adalah MA(1) dan AR(1), maka model ARIMA yang terbentuk adalah ARIMA(1,1,1). Tetapi tidak menutup kemungkinan terdapat model ARIMA lain yang terbentuk, maka perlu dilakukan proses trial and error untuk mengidentifikasi ordo- dan ordo- yang sesuai untuk mendapatkan model terbaiknya. Dari kriteria yang telah didapatkan diatas, model tentatif yang dapat terbentuk adalah ARIMA(2,1,1), ARIMA(1,1,2) dan ARIMA(2,1,2).



Gambar 3. Plot ACF dan PACF

Setelah didapatkan beberapa kombinasi dari ARIMA diatas, akan dipilih parameter yang signifikan. Pemilihan model ARIMA yang digunakan akan dilihat dari nilai *roots mean square error* (RMSE) terkecil. Dari hasil pengujian beberapa model ARIMA yang mungkin, diperoleh nilai RMSE untuk ARIMA(1,1,1) adalah 3.189418e+04, untuk ARIMA(2,1,1) adalah 2.866700e+04, nilai RMSE ARIMA(1,1,2) adalah 2.866700e+04 dan ARIMA(2,1,2) adalah 2.625190e+04. ARIMA (2,1,2) memiliki nilai RMSE terkecil, sehingga model yang dapat dipilih adalah ARIMA(2,1,2). ARIMA(2,1,2) menginterpretasikan model tersebut terdiri dari koefisien AR(1), AR(2), $d=1$, MA(1) serta MA(2). Persamaan model yang didapatkan adalah :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} - \theta_1 e_{t-1} - \theta_2 e_{t-2} + e_t$$

Tahap selanjutnya adalah verifikasi model ARIMA yang telah dipilih. Pemodelan deret waktu dengan ARIMA, harus memenuhi asumsi kenormalan sisaan, homoskedastisitas, dan keacakan sisaan (Enders 2004).

- Asumsi kenormalan sisaan
- Uji kenormalan sisaan dapat dilakukan dengan *JarqueBera Test*. Hipotesis uji ini adalah :
- H_0 = Residual berdistribusi normal
 H_1 = Residual tidak berdistribusi normal

(Ratnasari *et al* 2014)

```

Jarque-Bera normality test
-----
Population details:
  parameter of interest: skewness and kurtosis
  value under h_0:      "0 and 3"
  point estimate:       "-0.09052804276825316 and 2.0713575612398976"

Test summary:
  outcome with 95% confidence: fail to reject h_0
  one-sided p-value:           0.0853

Details:
  number of observations:     132
  JB statistic:               4.92337

```

Dengan bantuan paket *HypothesisTest.jl*, diperoleh hasil ujinya adalah gagal untuk menolak H_0 atau terima H_0 , maka residual berdistribusi normal. Akibatnya asumsi kenormalan sisaan terpenuhi.

- Asumsi kehomogenan ragam dapat dilihat dari plot ACF dan PACF. Terlihat bahwa asumsi kehomogenan ragam tidak terpenuhi.
- Asumsi kebebasan antar residual
Uji kebebasan antar sisaan dapat dilakukan dengan *Ljung-Box Test*. Hipotesis uji ini adalah :
 $H_0 = \rho_1 = \rho_2 = \dots = \rho_m = 0$ ((tidak ada korelasi residual antar lag).
 H_1 = Paling sedikit ada satu $\rho_k \neq 0$ dengan $k = 1, 2, \dots, m$ (ada korelasi residual antar lag)

(Ratnasari *et al* 2014)

```
Ljung-Box autocorrelation test
-----
Population details:
  parameter of interest:  autocorrelations up to lag k
  value under h_0:        "all zero"
  point estimate:         NaN

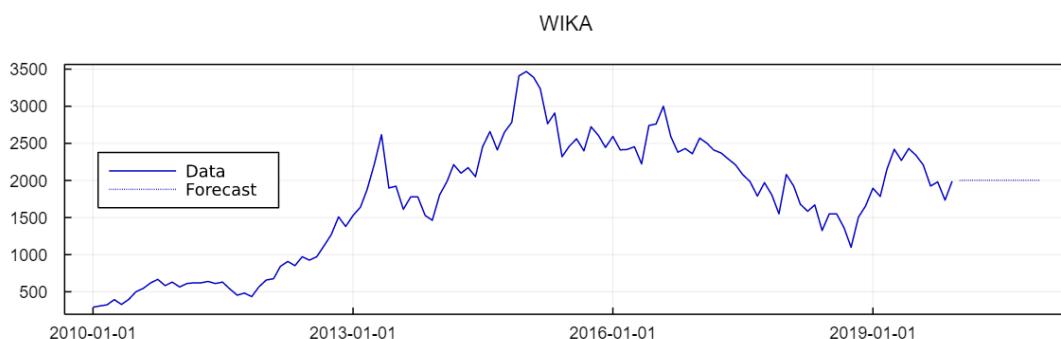
Test summary:
  outcome with 95% confidence: reject h_0
  one-sided p-value:           <1e-27

Details:
  number of observations:     132
  number of lags:            1
  degrees of freedom correction: 0
  Q statistic:               120.965
```

Dengan bantuan paket *HypothesisTest.jl*, diperoleh hasil ujinya adalah tolak H_0 , maka kesimpulannya ada korelasi residual antar lag. Akibatnya asumsi kebebasan antar sisaan tidak terpenuhi.

3.2 Peramalan Harga Saham

Metode yang digunakan untuk peramalan harga saham adalah metode ARIMA. Model ARIMA merupakan model peramalan yang tidak menggunakan pengaruh antar variabel seperti pada model regresi. Artinya, model ARIMA tidak memerlukan penjelasan mana variabel dependen dan independent (Nabila 2019). Hasil peramalan dengan model ARIMA(2,1,2) yang telah dipilih sebelumnya adalah tidak fit dengan data aktualnya, dimana data hasil *forecasting* menunjukkan bahwa harga saham cenderung konstan untuk 12 periode ke depan, sedangkan untuk data aktualnya mengalami fluktuasi. Hal ini dapat disebabkan karena ada kemungkinan model ARIMA yang dipilih bukan merupakan model ARIMA yang terbaik.



Gambar 4. Plot Hasil *forecasting*



Gambar 5. Plot data aktual vs data hasil *forecasting*

3.3 Model ARCH/GARCH

Hasil pengujian sebelumnya menunjukkan bahwa uji homoskedastisitas tidak terpenuhi. Hal ini dapat diatasi dengan memodelkan residual dengan menggunakan model ARCH-GARCH. Engle dalam Pindyck et al (1998) menyatakan bahwa ragam sisaan yang berubah-ubah terjadi karena ragam sisaan tidak hanya merupakan fungsi dari peubah bebas, akan tetapi tergantung dari seberapa besar sisaan dimasa lalu. Untuk mendeteksi apakah masih terdapat heteroskedastisitas dilakukan uji *Lagrange Multiplier* (LM).

Hipotesis LM:

$H_0 = \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$ ((tidak ada pengaruh ARCH dalam model).

H_1 = Paling sedikit ada satu $\alpha_k \neq 0$ dengan $k = 1, 2, \dots, m$ (ada pengaruh ARCH dalam model)

(Irawan 2019)

Banyaknya ordo ARCH yang dibentuk dapat dilihat pada seberapa banyak lag yang signifikan terhadap α pada pengujian LM. Ordo ARCH yang terlalu banyak akan menyebabkan model ragam yang terbentuk menjadi tidak efisien dan model ARCH lebih spesifik digunakan untuk model yang berordo ARCH rendah. Oleh karena itu digunakan perluasan dari model ARCH yaitu GARCH yang diperkenalkan oleh Tim Bollerslev pada tahun 1989 (Khoirunnisa 2014).

```
ARCH LM test for conditional heteroskedasticity
```

Population details:

parameter of interest:	T·R ² in auxiliary regression
value under h_0 :	0
point estimate:	106.264

Test summary:

outcome with 95% confidence:	reject h_0
p-value:	<1e-20

Details:

sample size:	120
number of lags:	5
LM statistic:	106.264

Hasilnya menunjukkan adanya pengaruh ARCH dalam model. Selanjutnya akan dilakukan pendugaan parameter GARCH dalam model. Diasumsikan model GARCH(1,1) adalah model yang dipilih.

```
GARCH{1, 1} model with Gaussian errors, T=113.
```

Mean equation parameters:

	Estimate	Std.Error	z value	Pr(> z)
μ	16.8992	12.1626	1.38944	0.1647

Volatility parameters:

	Estimate	Std.Error	z value	Pr(> z)
ω	730.216	625.785	1.16688	0.2433
β_1	0.634305	0.0873461	7.26198	<1e-12
α_1	0.457441	0.143877	3.17938	0.0015

Hasil pendugaan model GARCH(1,1) tersebut menunjukkan bahwa parameter β_1 dan α_1 signifikan. Hasil peramalan volatilitasnya untuk 12 periode ke depan adalah sebagai berikut.

```

12-element Vector{Float64}:
232.51885634328178
187.14681376172155
151.47960698668376
123.6326199973863
102.10575021304383
85.69257031941402
73.40337377339993
64.40408830844083
57.976213532741205
53.500170865984806
50.455595909932555
48.425291938586504

```

4. Simpulan

Harga saham yang berfluktuasi seiring berjalananya waktu, akan menyebabkan ragamnya terus meningkat pula. Kondisi tersebut ada kemungkinan dapat menyebabkan terjadinya heteroskedastisitas. Untuk mengatasi hal tersebut, salah satu model yang dapat digunakan adalah model ARCH dan GARCH yang berfungsi mengatasi variansi error yang tidak konstan (keragaman tinggi). Pada projek ini asumsi kenormalan ragam tidak terpenuhi sehingga mengindikasikan adanya efek ARCH dalam model. Ordo ARCH yang terlalu banyak akan menyebabkan model ragam yang terbentuk menjadi tidak efisien, sehingga model yang digunakan adalah model perluasannya yaitu model GARCH. Hasil evaluasi dan validasi model menunjukkan model yang dapat digunakan untuk data harga saham PT. WIKA Tbk. adalah model ARIMA(2,1,2)-GARCH(1,1). Namun hasil forecasting selama 12 periode ke depan menunjukkan hasil yang kurang fit dengan data asli. Hal ini dapat disebabkan karena model yang dipakai belum memenuhi kriteria model yang terbaik.

5. Daftar Pustaka

- Cryer JD, Chan KS. 2008. *Time Series Analysis, Second Edition*. USA: Springer.
- Enders, Walter. 2004. *Applied Econometric Time Series*. United States of America (US) : John Wiley & Sons
- Engle, Robert. 2001. The Use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives*.15(4):157-168.
- Irawan W. 2019. Peramalan harga saham PT. Univeler tbk. dengan menggunakan model arima. *Jurnal Matematika UNAND*. 4(3): 80-89. <http://jmua.fmpipa.unand.ac.id/index.php/jmua/article/viewFile/498/487>
- Kelikume I, Salami A. 2014. Time Series Modeling and Forecasting Inflation: Evidence From Nigeria. *The International Journal of Business and Finance Research*, Vol. 8 Ed. 2.
- Khoirunnisa E. 2014. Penerapan Metode ARCH/GARCH Pada Pemodelan Harga Penutupan Saham di Bursa Efek Indonesia Periode 2005-2013 [Skripsi]. Bogor: IPB University
- Montgomery DC, Jennings CL, Kulahci M. 2008. *Introduction to Time SeriesAnalysis and Forecasting*. Canada: John Wiley and Sons, Inc.
- Nabila M. 2019. Peramalan Harga Saham SM Entertainment Menggunakan Metode Single Exponential Smoothing dan Model ARIMA [Skripsi]. Bogor :IPB University.
- Puspitaningtyas Z. 2015. Prediksi Risiko Investasi Saham. Yogyakarta (ID): Pandiva Buku.
- Ratnasari D.H, Tarno, Yasin H.2014. Peramalan volatilitas menggunakan model *generalized autoregressive conditional heteroscedasticity in mean (garch-m)* (Studi Kasus pada Return Harga Saham PT. Wijaya Karya). *Jurnal Gaussian*. 3(4): <http://ejournal-s1.undip.ac.id/index.php/gaussian>
- Ratnasari, Dwi Hasti. 2014. Peramalan Volatilitas Menggunakan Model Generalized Autoregressive Conditional Heteroscedasticity in Mean (GARCH-M). UNDIP: Semarang.
- Ruey S. 2005. *Analysis of Time Series, Second Edition*. Wiley & Sons, Inc.
- Soeksin S.D, Fatolah C.2020. Peramalan Harga Saham PT. Bumi Serpong Damai Tbk. Dengan Metode Garch. *JAMAN (Jurnal Aplikasi Manajemen dan Akuntansi)*.1(1): <https://jurnal.poltekbisikal.ac.id/jaman/article/view/57>
- Seddighi H.R., K.A Lawler., A.V Katos. 2000. *Econometrics : A Practical Approach*. New York (US) : Taylor and Francis Group
- Tsay, R.S. 2002. *Analysis of Financial Time Series*. Canada: John Wiley and Sons, Inc.
- Zhang, G. P, Qi. M. 2005. Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2): 501–514. <https://doi.org/10.1016/j.ejor.2003.08.037>

Analisis Harga Saham WIKA Tahun 2010-2020 dengan Metode Simple Moving Average dan Exponential Moving Average

Fitra Nuvus Salsabila

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: fitrasalsabila@apps.ipb.ac.id

Abstrak

Bagi seorang investor maupun *trader*, sangat penting untuk melakukan analisis terhadap pergerakan harga saham. Hal ini dilakukan sebagai pertimbangan sebelum melakukan transaksi. Salah satu alat atau *tools* yang dapat digunakan untuk menganalisis saham adalah indikator. Indikator dapat membantu memperjelas kondisi pasar. *Moving Average* merupakan metode analisis yang dapat digunakan sebagai *trend indicator*. Sinyal indikasi yang ditunjukkan oleh *Moving Average* bersifat *lagging indicator*. Adapun *Moving Average* adalah pergerakan nilai rata-rata. Pada makalah ini, jenis metode *Moving Average* yang digunakan untuk melakukan analisis adalah *Simple Moving Average* (SMA) dan *Exponential Moving Average* (EMA). Metode ini diterapkan pada data *historical prices* saham WIKA, yang diambil harga *close*-nya setiap bulan selama tahun 2010-2020. Hasil analisis terhadap data tahun 2010-2019 digunakan untuk memperkirakan pergerakan saham tahun 2020. Berdasarkan hasil analisis, pada tahun 2010-2014, saham WIKA menunjukkan pergerakan *uptrend*. Setelah periode waktu tersebut, hingga tahun 2019 terjadi *downtrend*. Oleh karena itu, berdasarkan *trend* selama kurun waktu tersebut, dapat diperkirakan bahwa pergerakan pada tahun 2020 akan cenderung mengalami penurunan. Hasil analisis dengan menggunakan SMA12 dan EMA12 mengonfirmasi prediksi tersebut.

Kata Kunci: Saham, *Moving Average*, *Simple Moving Average*, *Exponential Moving Average*, Analisis Teknikal, Julia.

1. Pendahuluan

Investasi adalah “penanaman uang atau modal dalam suatu perusahaan atau proyek untuk tujuan memperoleh keuntungan” (KBBI, n.d.). Pada umumnya, masyarakat cenderung lebih memilih untuk menabung sebagai bentuk usaha mempersiapkan kebutuhan di masa depan. Namun, seiring berjalannya waktu, investasi semakin digandrungi masyarakat. Investasi dilakukan dengan membeli suatu barang, lalu menyimpannya dalam jangka waktu tertentu yang biasanya cukup panjang, hingga harga jual barang tersebut meningkat menjadi lebih tinggi daripada harga beli. Memang diperlukan modal awal untuk melakukan investasi. Namun, dengan melakukan investasi yang tepat, maka seseorang dapat memperoleh keuntungan setelah beberapa waktu (Haluansa, 2018). Salah satu produk investasi yang umum digunakan adalah saham.

Saham atau dikenal juga dengan istilah *stock*, merupakan instrumen pasar keuangan yang melibatkan perusahaan penerbit serta investor. Salah satu tujuan penerbitan saham oleh perusahaan adalah sebagai sumber pendanaan perusahaan. Tidak hanya perusahaan penerbit, para investor juga dapat memperoleh keuntungan dari transaksi saham sebagai investasi. “Saham dapat didefinisikan sebagai tanda penyertaan modal seseorang atau pihak (badan usaha) dalam suatu perusahaan atau perseroan terbatas. Dengan menyertakan modal tersebut, maka pihak tersebut memiliki klaim atas pendapatan perusahaan, klaim atas asset perusahaan, dan berhak hadir dalam Rapat Umum Pemegang Saham (RUPS)” (IDX, n.d.). Harga saham bersifat dinamis, yang artinya dapat berubah-ubah. Perubahan harga saham ini dapat disebabkan oleh berbagai faktor. Saham pada umumnya direpresentasikan dalam bentuk *candlestick*. *Candlestick* terdiri dari empat parameter, yakni harga pembukaan (*open*), harga tertinggi (*high*), harga terendah (*low*), dan harga penutup (*close*). *Candlestick* digunakan oleh para *trader* dan investor untuk memperkirakan pergerakan harga saham (Mitchell, 2021). Untuk menganalisis pergerakan harga saham, dapat digunakan metode analisis yang dibagi menjadi analisis fundamental dan analisis teknikal.

Analisis fundamental ialah analisis yang didasarkan pada kondisi suatu perusahaan, baik kondisi ekonomi dan industri terkait. Fundamental analysis pada umumnya menggunakan indikator-indikator perusahaan yang tertera melalui laporan keuangan perusahaan seperti *Price to Earning Ratio* (P/E), *Return to Equity* (ROE), dan lain-lainnya. Sementara itu, analisis teknikal ialah analisis yang didasarkan pada data-data harga historis yang terjadi pada pasar saham. Dalam penerapan analisis teknikal, prediksi untuk membeli atau menjual saham dilakukan dengan melihat grafik historis pergerakan saham (MOST, 2016).

Bagi seorang investor maupun *trader*, sangat penting untuk melakukan analisis terhadap pergerakan harga saham. Hal ini dilakukan sebagai pertimbangan sebelum melakukan transaksi. Tidak hanya memberikan keuntungan, investasi saham dapat juga memberikan risiko kerugian apabila *trader* maupun investor gagal membuat keputusan transaksi yang tepat. Salah satu alat atau *tools* yang dapat digunakan untuk menganalisis saham adalah indikator. Indikator dapat membantu memperjelas kondisi pasar. Indikator dibagi menjadi *leading indicator* dan *lagging indicator*. *Leading indicator* memberikan sinyal pergerakan sebelum harga saham itu benar-benar bergerak. Sedangkan *lagging indicator* memberikan

sinyal yang mengonfirmasi *behavior* dari pergerakan harga saham (Tradimo, n.d.). Terdapat banyak metode analisis yang dapat dijadikan indikator. Oleh karena itu, pada makalah ini, akan dibahas mengenai salah satu metode yang dapat digunakan untuk melakukan analisis harga saham. Adapun salah satu metode analisis yang dapat digunakan adalah *Moving Average*. *Moving Average* adalah *trend indicator* yang tergolong dalam analisis teknikal. "Trend merupakan arah kecenderungan pergerakan harga saham dalam periode waktu tertentu. Trend dibagi menjadi *uptrend* atau *bullish*, *downtrend* atau *bearish*, dan *sideways*" (POEMS, 2020). Pada kondisi *uptrend*, harga cenderung mengalami kenaikan. Pada kondisi *downtrend*, harga saham cenderung mengalami penurunan. Sedangkan pada kondisi *sideways*, tidak terjadi kenaikan dan penurunan yang kuat selama *range* waktu tertentu. Dengan menggunakan metode *Moving Average*, investor dan *trader* dapat memprediksi pergerakan harga saham.

Terdapat banyak jenis *Moving Average*, namun jenis yang akan dibahas di dalam makalah ini adalah *Simple Moving Average* dan *Exponential Moving Average*. Perbedaan antara kedua jenis *Moving Average* akan ditunjukkan dengan menganalisis kumpulan data berupa *historical prices* saham Wijaya Karya (Persero) Tbk per bulan selama Januari 2010 sampai dengan Desember 2020. Data selama tahun 2010-2019 akan dijadikan bahan analisis untuk memperkirakan *trend* di tahun 2020. Data tahun 2020 menjadi bukti untuk mengonfirmasi hasil dari analisis.

2. Data dan Metode

Metode analisis yang digunakan dalam makalah ini adalah *Moving Average*. *Moving Average* merupakan salah satu indikator dalam analisis teknikal. *Moving Average* sebagai indikator dapat digunakan baik secara mandiri, maupun dikombinasikan dengan indikator analisis teknikal lainnya (Cahyani & Mahyuni, 2020). *Moving Average* termasuk dalam kategori *trend indicator*, artinya *Moving Average* dapat digunakan untuk mengetahui *trend*. Terdapat berbagai jenis *Moving Average*. Penggunaan setiap metode *Moving Average* dapat dilakukan dengan cara yang sama. Perbedaan setiap metode *Moving Average* terdapat pada sensitivitasnya (Asthri, Topowijono, & Sulasmiyati, 2016). Apabila harga saham berada di atas garis *Moving Average*, kondisi *trend* adalah naik. Apabila harga saham berada di bawah garis *Moving Average*, kondisi *trend* adalah turun. Apabila garis *Moving Average* memotong harga dari atas, perubahan *trend* menuju naik. Sebaliknya, apabila garis *Moving Average* memotong harga dari bawah, perubahan *trend* menuju turun (Widodo & Hansun, 2015). Analisis ini dapat dilakukan terhadap berbagai periode, mulai dari periode waktu bulanan, mingguan, harian, bahkan menit. Melihat sudut garis hasil analisis *Moving Average* dapat menjadi salah satu cara untuk memanfaatkan metode ini. Apabila selama periode tertentu, dihasilkan pergerakan yang relatif horizontal, artinya harga tidak sedang mengalami *trend* yang kuat (Anderson, 2021). Dalam statistik, *Moving Average* adalah teknik analisis berupa penentuan titik data berdasarkan *subset* data dari *dataset* lengkapnya. *Moving Average* dibagi menjadi *Simple Moving Average* (SMA) dan *Exponential Moving Average* (EMA). Pada makalah ini, *dataset* yang akan dianalisis adalah *historical prices* saham Wijaya Karya (Persero) Tbk, atau WIKA. *Dataset* yang digunakan merupakan kumpulan data berupa *historical prices* saham WIKA per bulan pada tahun 2010-2020.

2.1. Alat dan Bahan

Alat dan bahan yang digunakan adalah sebagai berikut:

Tabel 1. Alat dan bahan

Alat dan Bahan	Keterangan
Personal Computer (PC)	Intel(R) Core(TM) i5-8265U, CPU @ 1.60GHz, 1.80 GHz, (3.89 GB usable), 64-bit operating system, x64-based processor
Software Julia	Versi 1.5.1
CSV.jl	Paket Julia untuk mengolah <i>delimited files</i>
DataFrames.jl	Paket Julia untuk mengolah data tabular
Plots.jl	Paket Julia untuk visualisasi data
StatsPlots.jl	Paket Julia untuk visualisasi data
PlotlyJS.jl	Paket Julia untuk <i>backend</i> visualisasi data
Measures.jl	Paket Julia untuk mendefinisikan satuan pengukuran
StatsBase.jl	Paket Julia untuk menyediakan berbagai fungsi statistik
MarketTechnicals.jl	Paket Julia untuk menyediakan berbagai fungsi umum dalam analisis teknikal

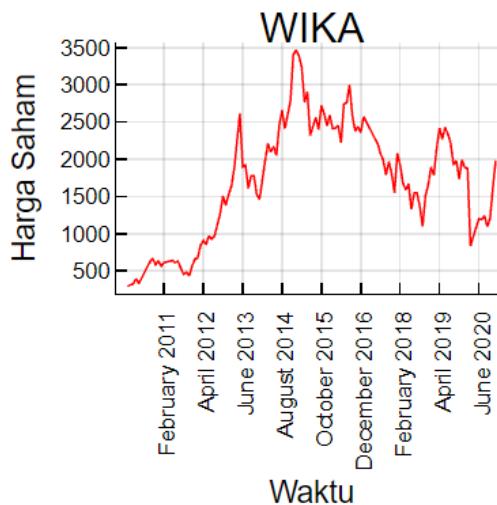
2.2. Data

Data *historical prices* saham WIKA diperoleh dari situs Yahoo Finance. Tabel 2 menunjukkan deskripsi statistik data harga *close* saham WIKA selama periode tahun 2010-2020. Adapun deskripsi statistik data meliputi jumlah data (*length*), jumlah data yang hilang (*missing count*), rata-rata (*mean*), nilai terkecil (*minimum*), kuartil 1, kuartil 2 atau nilai tengahnya (*median*), kuartil 3, dan nilai terbesarnya (*maximum*).

Tabel 2. Deskripsi Statistik Harga Saham WIKA 2010-2020

Parameter	Deskripsi
<i>Length</i>	132
<i>Missing Count</i>	0
<i>Mean</i>	1700.454545
<i>Minimum</i>	292.000000
<i>1st Quartile</i>	972.000000
<i>Median</i>	1797.500000
<i>3rd Quartile</i>	2372.500000
<i>Maximum</i>	3468.000000
<i>Type</i>	Int64

Grafik pada gambar 1 di bawah ini menunjukkan analisis pergerakan harga saham WIKA per bulan selama periode tahun 2010-2020. Berdasarkan grafik, dapat diketahui bahwa sejak tahun 2010 sampai dengan 2014, harga saham WIKA relatif meningkat, dengan puncaknya pada periode antara Agustus 2014 sampai Oktober 2015.

**Gambar 1.** Pergerakan harga saham WIKA selama tahun 2010-2020

Agar dapat dilihat dengan lebih jelas perbandingan harga saham WIKA dari tahun ke tahun selama periode tahun 2010-2020, peneliti menjumlahkan harga saham setiap bulan dalam satu tahun selama periode waktu tersebut, kemudian membagi hasilnya dengan jumlah bulan selama setahun, yakni 12, sehingga diperoleh rata-rata harga saham per tahun. Adapun data harga rata-rata saham WIKA selama tahun 2010-2020 dapat dilihat pada tabel 3.

Tabel 3. Rata-Rata Harga Saham WIKA Tahun 2010-2020

Tahun	Rata-rata harga (Rp)
2010	466
2011	564
2012	1007
2013	1887
2014	2391
2015	2773
2016	2530
2017	2128
2018	1583
2019	2094
2020	1348

Melalui *bar chart* pada gambar 2, yang didasarkan pada data rata-rata harga saham WIKA per tahun, dapat dianalisis perbandingan rata-rata harga saham WIKA selama periode tahun 2010-2020. Pada periode tahun 2010-2014, rata-rata

harga saham WIKA cenderung meningkat. Puncak rata-rata harga saham WIKA terjadi pada tahun 2015. Pada periode tahun 2016-2020 terjadi penurunan rata-rata harga saham WIKA jika dibandingkan dengan tahun-tahun sebelumnya.



Gambar 2. Rata-rata harga saham WIKA tahun 2010-2020

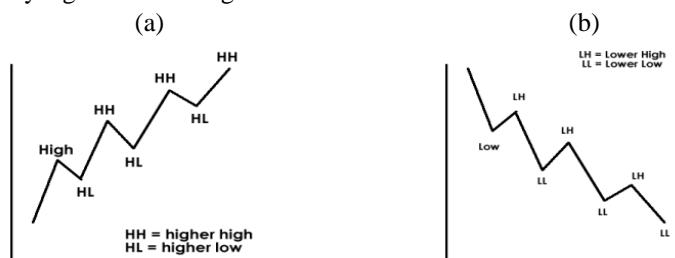
Gambar 3 di bawah ini merupakan grafik *candlestick* dari *historical prices* saham WIKA pada tahun 2010-2020 dilihat dari harga ketika *close* per bulannya. Pada data berikut, diterapkan metode *Simple Moving Average* dan *Exponential Moving Average* untuk memperkirakan kondisi pergerakan saham WIKA.



Gambar 3. Candlestick historical prices saham WIKA per bulan pada tahun 2010-2020

2.3 Trend Saham

Pada gambar 4 dapat dilihat pola suatu pergerakan dapat dikategorikan *uptrend* atau *downtrend*. Di dalam suatu *trendline* terdapat dua parameter utama berupa puncak, yang dikenal juga sebagai *peak* atau *high*, serta lembah, yang dikenal juga *trough* atau *low*. Suatu *trendline* dapat dikatakan *uptrend* apabila memiliki deretan lembah dan puncak yang cenderung naik. Meskipun dalam kondisi *uptrend* dapat terbentuk beberapa lembah, jika dilihat dari periode waktu tertentu yang lebih panjang, secara keseluruhan *trendline* yang terbentuk berupa kenaikan. Adapun suatu *trendline* dapat dikatakan *downtrend* apabila terbentuk deretan puncak dan lembah yang cenderung menurun. Sedangkan, pada kondisi *sideways*, data biasanya memiliki pola *high* dan *low* yang terus berulang selama kurun waktu tertentu. Pada kondisi *sideways*, tidak disarankan untuk melakukan transaksi (Edusaham, 2020). Ilustrasi pada gambar 4 yang diperoleh dari Real Fibonacci Trading (2013), menggambarkan pola *low-high-higher low-higher high-higher low-higher high* untuk *uptrend*. *Downtrend* memiliki pola *high-low-lower high-lower low-lower high-lower low*. Sedangkan pada kondisi *sideways*, data biasanya memiliki pola *high* dan *low* yang terus berulang selama kurun waktu tertentu.



Gambar 4. (a) pola *uptrend*, (b) pola *downtrend*

2.4. Simple Moving Average

Simple Moving Average merupakan analisis *Moving Average* yang paling sederhana. Analisis *Simple Moving Average* tidak menerapkan pembobotan pada harga saham. Artinya, setiap harga memiliki bobot yang sama (Widodo & Hansun, 2015). Dalam analisis harga saham, *Simple Moving Average* adalah perhitungan yang mengambil rata-rata dari sekumpulan harga selama periode waktu tertentu di masa lalu. Periode waktu yang umum digunakan dalam analisis harga saham dengan metode ini adalah 15, 20, 30, 100, atau 200 hari sebelumnya. Semakin pendek rentang waktu yang diambil dan dirata-rata, maka titik data berupa rata-rata yang diperoleh akan semakin sensitif terhadap perubahan harga. Semakin lama rentang waktu, semakin tidak sensitif rata-ratanya. Periode waktu yang lebih pendek biasanya digunakan oleh *trader*, sedangkan periode waktu yang lebih panjang biasanya digunakan oleh investor (FOREXimf, n.d.).

SMA dapat diperoleh dengan menerapkan persamaan berikut:

$$SMA = \frac{\sum_i^n P_i}{n} \quad (1)$$

Keterangan:

- | | |
|-------|----------------------|
| P_i | = Price/ Harga saham |
| n | = Periode waktu |

2.5. Exponential Moving Average

Perbedaan *Exponential Moving Average* dengan *Simple Moving Average* ada pada pemberian bobotnya. *Exponential Moving Average* adalah jenis *Moving Average* yang menempatkan bobot lebih besar pada titik data terbaru. Sementara *Simple Moving Average* memberikan bobot yang sama untuk semua nilai. Untuk menggunakan *Exponential Moving Average*, diperlukan observasi lebih dibandingkan dengan menggunakan *Simple Moving Average*. Sebagai contoh, apabila dilakukan observasi selama periode waktu 20 hari lalu, maka perlu dihitung terlebih dahulu *Simple Moving Average* dari periode 20 hari tersebut. Pada hari ke-21, nilai *Simple Moving Average* yang sebelumnya diperoleh akan digunakan sebagai *Exponential Moving Average* hari kemarin. Selanjutnya, berdasarkan periode waktu yang diobservasi dan *smoothing value*, akan ditentukan sebuah pengali. *Smoothing value* yang umum digunakan adalah 2 (Chen, 2021).

Adapun, secara matematis, rumus untuk menentukan pengali dapat dituliskan sebagai berikut:

$$k = \frac{s}{n+1} = \frac{2}{n+1} \quad (2)$$

Keterangan:

- | | |
|-----|-------------------|
| k | = Nilai pengali |
| s | = Smoothing value |
| n | = Periode waktu |

Setelah memperoleh nilai pengali, maka dapat ditentukan EMA hari ini, yang secara matematis dapat dirumuskan sebagai berikut:

$$EMA_t = (k \times P_t) + ((1 - k) \times EMA_{t-1}) \quad (3)$$

Keterangan:

- | | |
|-------------|-----------------------------|
| EMA_t | = Nilai EMA hari ini |
| EMA_{t-1} | = Nilai EMA hari sebelumnya |
| P_t | = Harga saham hari ini |

3. Hasil dan Pembahasan

Diambil harga *closing* saham WIKA per bulan selama tahun 2010-2020 untuk di-*plot* dan dilakukan analisis dengan metode *Simple Moving Average* dan *Exponential Moving Average*. Karena *dataset* ini merupakan kumpulan data perbulan selama tahun 2010-2020, maka terdapat 12 *candlestick* yang terbentuk di setiap satu tahun. Oleh karena itu, dilakukan analisis *Simple Moving Average* dan *Exponential Moving Average* dengan nilai n atau periodonya 12, atau dapat disebut juga SMA12 dan EMA12. Pada analisis ini, diasumsikan bahwa data selama tahun 2020 adalah data yang akan diprediksi dengan didasarkan pada analisis *trend* tahun 2010-2019.

3.1 Simple Moving Average

Tabel 4 menunjukkan titik data yang diperoleh dari perhitungan SMA12. Melalui perhitungan tersebut, diperoleh 110 titik data.

Tabel 4. Titik data SMA12

Titik Data				
466.124	849.672	1998.775	2502.713	1912.500
488.890	939.193	2051.638	2527.795	1860.833
513.971	1007.105	2124.181	2564.452	1787.083
538.666	1079.647	2228.750	2580.658	1742.917
557.573	1159.907	2390.813	2552.104	1706.667
583.426	1245.955	2529.338	2537.363	1671.250
601.176	1355.540	2646.255	2530.293	1598.750
611.980	1502.554	2731.531	2528.376	1573.750
611.208	1579.727	2787.095	2535.674	1582.500
597.317	1662.688	2848.447	2535.087	1567.083
581.883	1715.937	2870.828	2528.079	1555.417
569.535	1770.730	2871.213	2533.697	1594.583
564.133	1813.175	2863.110	2489.433	1664.167
571.850	1814.718	2861.953	2432.791	1714.167
577.252	1821.664	2868.127	2348.168	1806.250
595.774	1844.815	2853.464	2281.251	1872.083
619.697	1873.755	2773.204	2247.083	1927.083
637.447	1901.923	2700.276	2195.000	1973.750
667.545	1891.505	2618.858	2127.500	2047.083
692.240	1854.462	2550.561	2086.667	2066.250
728.511	1867.196	2524.708	2038.750	2094.167
784.075	1911.570	2467.600	1977.917	2093.750

Berdasarkan grafik di bawah ini, garis analisis SMA12 menunjukkan bahwa *trend* data adalah *uptrend* pada tahun 2010-2014. Namun, *trend* mulai menurun setelah periode tersebut. Dilihat dari *trend* pada Juni 2014-April 2019, dimulai dari sekitar bulan Juni 2014, pada sejumlah titik, diperoleh pola titik data *high-low-lower high-low-lower low-low-high*. Pola tersebut menunjukkan kondisi *downtrend*. Kondisi ini dikonfirmasi oleh grafik pergerakan harga saham berada di bawah garis SMA12, yang artinya kondisi sedang menurun. Pada periode sekitar tahun 2019, harga saham mengalami peningkatan, meskipun secara keseluruhan masih dalam kondisi *downtrend*. Namun, karena garis SMA12 memotong garis harga dari bawah, maka peralihan *trend* yang terjadi adalah menuju kondisi turun. Berdasarkan analisis, dapat diperkirakan pada tahun selanjutnya akan terjadi *downtrend*.



Gambar 3. Grafik analisis *historical prices* saham WIKA dengan SMA12

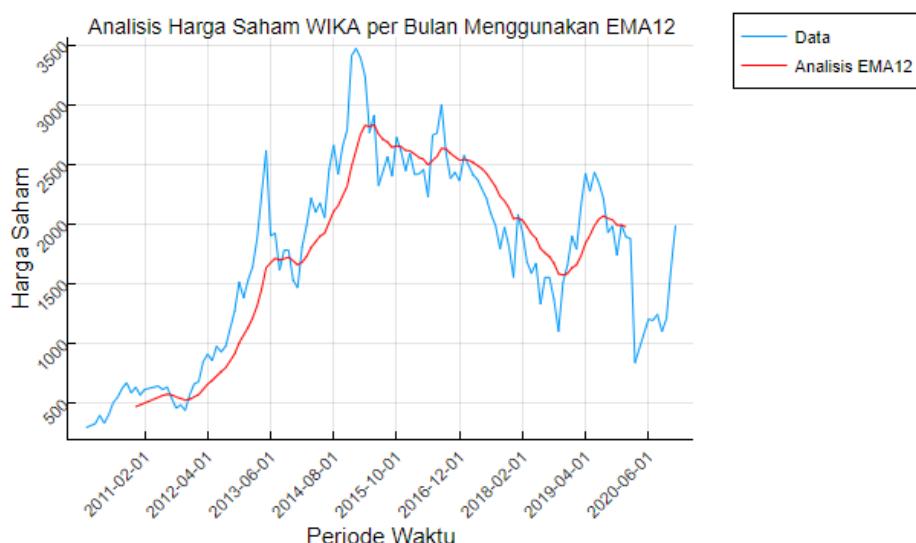
3.2 Exponential Moving Average

Tabel 5 menunjukkan titik data yang diperoleh dari perhitungan EMA12. Melalui perhitungan tersebut, diperoleh 110 titik data.

Tabel 5. Titik data EMA12

Titik Data				
466.124	908.874	2103.109	2529.581	1914.976
481.321	1001.278	2150.696	2564.984	1877.287
501.303	1059.520	2227.292	2631.983	1792.320
519.637	1131.597	2312.762	2625.987	1755.040
535.149	1209.682	2481.253	2588.144	1723.495
551.125	1312.085	2633.082	2563.814	1668.342
560.369	1452.160	2749.443	2532.458	1580.905
571.040	1631.237	2824.395	2538.234	1569.227
565.822	1672.346	2815.154	2532.352	1582.423
548.584	1710.694	2829.418	2513.528	1630.512
538.272	1695.413	2751.017	2491.447	1654.279
522.423	1708.128	2706.049	2460.455	1730.544
528.959	1718.887	2683.671	2421.924	1836.614
548.736	1689.523	2639.803	2369.320	1903.289
568.320	1654.703	2652.550	2310.194	1984.321
610.537	1677.956	2645.526	2230.164	2039.041
656.231	1725.413	2614.651	2190.139	2065.343
686.348	1800.475	2611.321	2130.887	2043.751
730.352	1846.180	2580.721	2041.519	2033.943
760.463	1896.251	2555.541	2047.440	1987.952
793.065	1920.097	2539.934	2028.603	1988.267
843.447	2002.251	2491.109	1974.972	1973.149

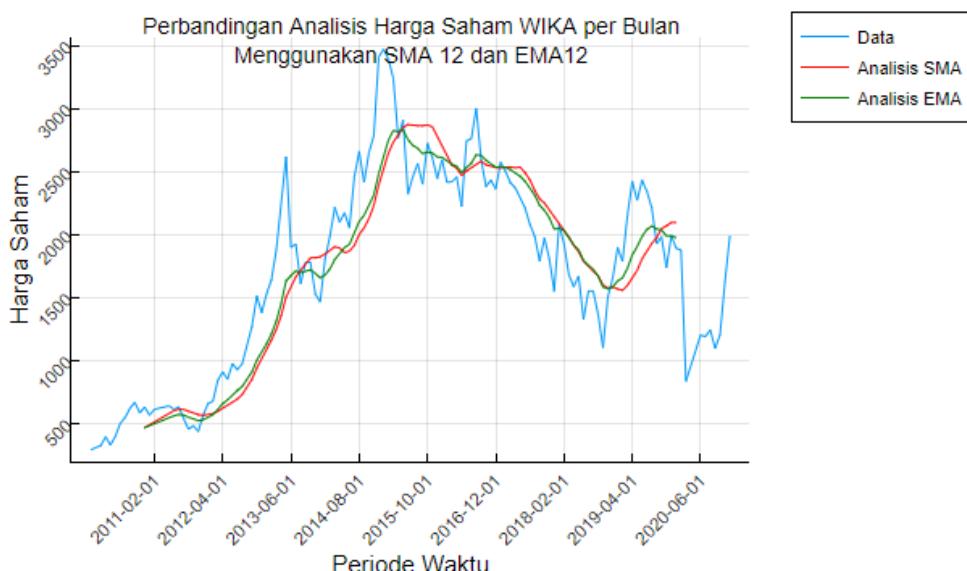
Berdasarkan garis *plot* analisis EMA12, dapat dilihat bahwa terjadi kondisi *uptrend* pada tahun 2010 sampai dengan sekitar tahun 2014. Dimulai dari sekitar bulan Juni 2014, pada sejumlah titik, diperoleh pola titik data *high-low-lower high-low low-lower high*. Pola tersebut menunjukkan kondisi *downtrend*. Kondisi ini dikonfirmasi oleh grafik pergerakan harga saham berada di bawah garis EMA12, yang artinya kondisi sedang menurun. Pada periode sekitar tahun 2019, harga saham mengalami peningkatan, meskipun secara keseluruhan masih dalam kondisi *downtrend*. Namun, karena garis EMA12 memotong garis harga dari bawah, maka peralihan *trend* yang terjadi adalah menuju kondisi turun. Berdasarkan *trend* pada periode mulai dari Juni 2014 hingga sekitar tahun 2019, dapat diprediksi akan terjadi *bearish* pada periode berikutnya, yakni tahun 2020.



Gambar 4 Analisis historical prices WIKA dengan EMA12

3.3 Perbandingan Simple Moving Average dan Exponential Moving Average

Grafik pada gambar 7 di bawah ini menunjukkan perbandingan hasil analisis dengan menggunakan *Simple Moving Average* dan *Exponential Moving Average*. Data yang digunakan merupakan data yang sama dengan analisis pada sub bab 3.1 dan 3.2, yakni data *historical prices* saham WIKA per bulan pada tahun 2010-2020. Berdasarkan grafik, dapat dilihat bahwa hasil analisis dengan *Exponential Moving Average* apabila dibandingkan dengan hasil analisis dengan *Simple Moving Average*, cenderung lebih sensitif terhadap perubahan harga saham. Apabila harga saham mengalami perubahan, *Exponential Moving Average* lebih dahulu mengalami perubahan pergerakan, dibandingkan dengan *Simple Moving Average*. Hal ini karena *Exponential Moving Average* memberikan bobot yang lebih besar pada data baru. Sedangkan *Simple Moving Average* memberikan bobot yang sama untuk semua data.



Gambar 5. Perbandingan analisis *historical prices* saham WIKA dengan SMA12 dan EMA12

4. Simpulan

Berdasarkan hasil analisis, dapat ditarik kesimpulan bahwa, ketika kedua metode, yakni *Simple Moving Average* dan *Exponential Moving Average*, diterapkan pada data *historical prices* saham WIKA pada tahun 2010-2019 untuk memperkirakan pergerakan saham WIKA pada tahun 2020, dapat diprediksi pada tahun 2020 akan terjadi *bearish* atau penurunan. Selain itu, meskipun menggunakan periode waktu (n) yang sama, hasil analisis dengan *Exponential Moving Average* cenderung lebih sensitif terhadap perubahan harga saham. Hal ini disebabkan oleh perbedaan pemberian bobot data oleh kedua metode tersebut.

5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Rahmawati Masithoh selaku rekan yang telah membantu penulis dalam proses pengajaran makalah ini.

6. Daftar Pustaka

- Anderson, S. (24 Agustus 2021). *Trend Trading: The 4 Most Common Indicators*. [diakses 17 Oktober 2021].
Investopedia: <https://www.investopedia.com/articles/active-trading/041814/four-most-commonly-used-indicators-trend-trading.asp>
- Asthri, D. D.;Topowijono; & Sulasmiyati, S. (2 April 2016). Analisis Teknikal dengan Indikator Moving Average Convergence Divergence untuk Menentukan Sinyal Membeli dan Menjual dalam Perdagangan Saham (Studi pada Perusahaan Sub Sekto Makanan dan Minuman di Bei Tahun 2013-2015). *Jurnal Administrasi Bisnis*.
- Cahyani, N. N.;& Mahyuni, L. P. (2020). Akurasi Moving Average dalam Prediksi Saham LQ45 di Bursa Efek Indonesia.
- Chen, J. (28 April 2021). *Exponential Moving Average (EMA)*. (J. Chen, Tuottaja) [diakses 9 Oktober 2021].
Investopedia: <https://www.investopedia.com/terms/e/ema.asp>
- Edusaham. (4 Mei 2020). *Apa Itu Trendline dalam Saham (Uptrend, Downtrend, dan Sideways)*. [diakses 24 Oktober 2021]. Edusaham: <https://www.edusaham.com/2019/03/apa-itu-trendline-dalam-saham-up-trend-down-trend-sideways.html>

- FOREXimf. *Moving Average*. [diakses 7 Oktober 2021]. foreximf: <https://www.foreximf.com/course/kelas-dasar/moving-average>
- Haluansa, N. A. (2018). *Analisis Fundamental dan Teknikal untuk Menentukan Keputusan Investasi Saham Sub Sektor Food and Beverages*. Semarang.
- IDX. *Saham*. [diakses 7 Oktober 2021]. IDX: <https://www.idx.co.id/produk/saham/>
- KBBI. [diakses 23 Oktober 2021]. Kamus Besar Bahasa Indonesia: <https://kbbi.web.id/investasi>
- Mitchell, C. (28. April 2021). *How to Use a Moving Average to Buy Stocks*. [diakses 17 Oktober 2021]. Investopedia: <https://www.investopedia.com/articles/active-trading/052014/how-use-moving-average-buy-stocks.asp>
- Mitchell, C. (25 Maret 2021). *Understanding Basic Candlestick Charts*. [diakses 17 Oktober 2021]. Investopedia: <https://www.investopedia.com/trading/candlestick-charting-what-is-it/>
- MOST. (6 Januari 2016). *Yuk Belajar Main Saham : Perbedaan Analisa Fundamental dan Teknikal*. [diakses 7 Oktober 2021]. Mandiri Online Securities Trading: <https://www.most.co.id/tips-investasi/yuk-belajar-main-saham-perbedaan-analisa-fundamental-dan-teknikal>
- POEMS. (Juli 2020). *Memulai Analisis Teknikal Pertama Anda*. [diakses 17 Oktober 2021]. Philip Sekuritas Indonesia: https://www.poems.co.id/htm/Freducation/LPNewsletter/v87/vol87_analisisteknikal.html
- Real Fibonacci Trading. (28 April 2013). *Higher High? Higher Low? Lower Low? Lower High?*. [diakses 18 Oktober 2021]. Real Fibonacci Trading: <http://www.realfibonacciitrading.com/2013/04/28/higher-high-higher-low-lower-low-lower-high/>
- Tradimo. *How to Work with Indicator*. [diakses 17. Oktober 2021]. Tradimo: <https://learn.tradimo.com/technical-analysis-how-to-work-with-indicators/overview>
- Widodo, D.;& Hansun, S. (2015). *Implementasi Simple Moving Average dan Exponential Moving Average dalam Menentukan Tren Harga Saham Perusahaan*. Tangerang.

Analisis Data Saham Menggunakan Metode K-Means Clustering dengan Visualisasi Data dalam Bentuk Plot/Scatter dan Boxplot

Yohanes Mario Defianus Beti

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: yohanesmario@apps.ipb.ac.id

Abstrak.

Dalam analisis saham, ada dua jenis analisis yang sering digunakan yaitu analisis fundamental saham dan analisis teknikal. Analisis fundamental adalah teknik analisis saham yang menitikberatkan beberapa faktor seperti, kinerja perusahaan yang bersangkutan, persaingan usaha, industri, hingga kondisi ekonomi (makro dan mikro). Intinya, lewat analisis ini, akan terlihat jelas apakah saham dari sebuah perusahaan dinyatakan sehat atau tidak.. Analisis teknikal adalah teknis analisis yang ditujukan untuk mengamati pola-pola seperti data pasar, harga saham, dan volume transaksi saham. Artinya, bukan melihat hal-hal yang berkaitan dengan kondisi ekonomi, kesehatan keuangan perusahaan, serta prospek bisnis ke depan. Pada penulisan makalah ini, penulis menggunakan analisis teknikal pada asumsi dasar harga saham dan return saham. Penulis menganalisis data saham 4 saham perusahaan di Indonesia menggunakan metode *K-means Clustering* dengan visualisasi data dalam *plot/scatter* dan *boxlot*. K-Means merupakan salah satu metode pengelompokan data non-hirarki yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini bertujuan mempartisi data yang ada ke dalam kelompok sehingga data berkarakteristik sama dimasukan kedalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokan ke dalam kelompok yang lain. Nilai signifikansi menunjukkan terdapat perbedaan yang berarti antara cluster1, cluster 2 dan cluster 3 yang berhubungan dengan atribut-atribut data saham.

Kata Kunci: Saham, *K-Means*, *Clustering*.

1. Pendahuluan

Saham adalah sebuah bukti kepemilikan nilai sebuah perusahaan. Artinya pemilik saham adalah pemilik perusahaan. Semakin besar saham yang dimiliki, maka semakin besar kekuasaannya di perusahaan tersebut. Dengan menerbitkan saham, memungkinkan perusahaan-perusahaan yang membutuhkan pendanaan jangka panjang untuk menjual kepentingan dalam bisnis - saham (efek ekuitas) -dengan imbalan uang tunai. Ini adalah metode utama untuk meningkatkan modal bisnis selain menerbitkan obligasi. Saham dijual melalui pasar primer (primary market) atau pasar sekunder (secondary market). Saham (stock) merupakan salah satu instrumen pasar keuangan yang paling popular. Menerbitkan saham merupakan salah satu pilihan perusahaan ketika memutuskan untuk pendanaan perusahaan. Pada sisi lain, saham merupakan instrumen investasi yang banyak dipilih oleh para investor karena saham mampu memberikan tingkat keuntungan yang menarik. Saham dapat didefinisikan sebagai tanda penyertaan modal seseorang atau pihak (badan usaha) dalam suatu perusahaan atau perseroan terbatas. Dengan menyertakan modal tersebut, maka pihak tersebut memiliki klaim atas pendapatan perusahaan, klaim atas aset perusahaan, dan berhak hadir dalam Rapat Umum Pemegang Saham (RUPS).

Dalam analisis saham, ada dua jenis analisis yang sering digunakan yaitu analisis fundamental saham dan analisis teknikal. Analisis fundamental adalah teknik analisis saham yang menitikberatkan beberapa faktor seperti, kinerja perusahaan yang bersangkutan, persaingan usaha, industri, hingga kondisi ekonomi (makro dan mikro). Intinya, lewat analisis ini, akan terlihat jelas apakah saham dari sebuah perusahaan dinyatakan sehat atau tidak. Data-data dari analisis tersebut akan digunakan sebagai acuan untuk membeli saham.

Selain analisis fundamental, ada juga analisis teknikal saham. Analisis teknikal adalah teknis analisis yang ditujukan untuk mengamati pola-pola seperti data pasar, harga saham, dan volume transaksi saham. Artinya, bukan melihat hal-hal yang berkaitan dengan kondisi ekonomi, kesehatan keuangan perusahaan, serta prospek bisnis ke depan.

Pada penulisan makalah ini, penulis ingin menganalisis data saham 4 saham perusahaan di Indonesia menggunakan analisis teknikal pada asumsi dasar harga saham dan return saham. Penulis menganalisis data saham 4 saham perusahaan di Indonesia menggunakan metode *K-means Clustering* dengan visualisasi data dalam *plot/scatter* dan *boxlot*. K-Means merupakan salah satu metode pengelompokan data non-hirarki yang berusaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini akan mempartisi data yang ada ke dalam kelompok sehingga data berkarakteristik sama dimasukan kedalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokan kedalam kelompok yang lain. Nilai signifikansi menunjukkan terdapat perbedaan yang berarti antara cluster 1, cluster 2 dan cluster 3 yang berhubungan dengan atribut-atribut pada penelitian.

2. Data dan Metode

Data dan metode yang digunakan dalam penulisan makalah ini, meliputi :

2.1. Alat dan Bahan

Alat pendukung pengerjaan tugas ini dikerjakan di laptop Acer Swift 3x (SF314-510G) dengan spesifikasi sebagai berikut : 1) prosesor Intel® Core™ i5-1135G7 processor, 2) OS : Windows 10 Home, 3) Memory RAM : 16GB LPDDR4X memory, 4) Storage : 512 GB PCIe Gen3, 8 Gb/s, NVMe, 5) Inch, Res, Ratio, Panel : 14.0" IPS Full HD (1920 x 1080), 300 nits, Acer ComfyView™ LED-backlit TFT LCD, 100% sRGB, Wide viewing angle up to 170 degree, 6) Graphics : Intel® Iris® Xe Max Graphics (4GB VRAM), 7) Features : Iris Xe Max, Thunderbolt 4™, Fingerprint, Backlight

Software yang digunakan yaitu Jupyter Notebook (Julia versi 1.6.2 dan Python 3.0). Paket yang digunakan meliputi : 1) Paket XLXS. Paket ini digunakan untuk membaca data excel pada jupyter. 2) Paket DataFrames adalah struktur datayang memiliki kolom berlabel, yang secara individu memiliki tipe data berbeda, Seperti SQL atau speadsheet. Paket ini menyediakan tiga tipe data meliputi NA yakni nilai yang hilang di julia diwakili oleh tipe data tertentu, Dataarray yakni fungsi khusu untuk menganalisis data, DataFrame yakni struktur data 2-D seperti spreadsheet. 3) Paket Plot, digunakan untuk menvisualisasi data dalam bentuk 2 dimensi. 4) StatsBase, digunakan untuk mendeskripsikan statistika(mean, kuartil, maksimum minimum dll) harga saham dan retrun saham. 5) DecisionTree, digunakan untuk menganalisis data saham dan return saham, 6) RDataset, Paket ini pada dasarnya adalah port sederhana yang dengan mudah mengumpulkan kumpulan data dari banyak paket R standar di satu lokasi yang nyaman.

2.2. Data

Data yang digunakan dalam analisis ini adalah data saham empat emiten di Indonesia tahun 2010 sampai 2020 meliputi: Wijaya Karya (WIKA) Tbk., BISI International Tbk., Bumi Serpong Damai (BSDE) Tbk., dan Pelat Timah Nusantara (NIKL) Tbk. Serta jenis datanya adalah data sekunder dengan 2 *feature*-nya adalah harga saham dan *return* saham.

2.3. Metode K-Means Clustering

Clustering adalah pengelompokan data berdasarkan kriteria, jarak, karakteristik/ ciri dan kemiripan. Pendekatan dasar pembuatan cluster dimulai dari pengelompokan objek data. Pengelompokan ini seharusnya hanya terjadi berdasarkan data yang menggambarkan objek. Objek serupa atau menunjukkan hubungan yang sama dikelompokkan bersama. Objek yang berbeda disimpan di cluster lain. Visualisasi dalam bentuk plot line/scatter dengan jelas menunjukkan beberapa cluster berbeda yang terbentuk ketika ada lebih banyak kesamaan antara objek data yang berbeda dalam sebuah cluster dan perbedaan dengan objek data dari cluster lain.

Secara umum, teknik pengklasteran dapat diklasifikasikan ke dalam dua metode, yaitu *Hierarchical Method* dan *Non-Hierarchical Method* (Hair, dkk. 1998). Pada penulisan makalah ini peneliti menggunakan metode *Non-Hierarchical Method* yakni *K-means Clustering*. K-means merupakan salah satu algoritma clustering. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas. Pada algoritma pembelajaran ini, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap cluster terdapat titik pusat (centroid) yang merepresentasikan klaster tersebut.

Memahami algoritma K-means akan memberikan pandangan yang lebih baik tentang bagaimana mendekati masalah. Berikut ini langkah-langkah algoritma K-means:

- 1) Menentukan klaster (K) $x_1, \dots, x_n, \forall n \in \mathbb{N}$
Penulisan dalam Julia sebagai berikut;
`km = KMeans(n_clusters=6)`
- 2) Pilih sejumlah centroid $c_1, \dots, c_k, k \in K$ tempatkan di tempat yang random.
Rumus umum penentuan klaster
- 3) Tetapkan titik data ke centroid terdekat. Langkah ini akan membentuk klaster.

Penentuan jarak titik data ke centroid menggunakan metode *Euclidean Distance*

$$D(a, b) = \sqrt{\sum_{k=1}^n (a_k - b_k)^2}$$

di mana

n adalah jumlah dimensi(atribut)

a_k dan b_k atribut ke-k dari objek data

Dalam Julia dapat ditulis sebagai berikut :

```
y_predicted = km.fit_predict(data2[['Nama emiten']])
y_predicted

data1 = data[data.cluster==0]
data2 = data[data.cluster==1]
data3 = data[data.cluster==2]
data4 = data[data.cluster==3]
data5 = data[data.cluster==4]
data6 = data[data.cluster==5]

plt.scatter(data1.Waktu, data1['emiten1'], s=10, color='green')
plt.scatter(data2.Waktu, data2['emiten1'], s=10, color='red')
plt.scatter(data3.Waktu, data3['emiten1'], s=10, color='yellow')
plt.scatter(data4.Waktu, data4['emiten1'], s=10, color='black')
plt.scatter(data5.Waktu, data5['emiten1'], s=10, color='pink')

km.cluster_centers_
```

- 4) Menentukan centroid yang baru/rata-rata data yang ada di masing-masing klaster
- 5) Ulangi langkah 3, apabila masih ada data yang berpindah kalster ata ada perubahan nilai centroid. Jika tidak ada perubahan pada klaster, maka hentikan proses klaster.

Data emiten yang berkarakteristik sama dimasukan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokan kedalam kelompok yang lain. Nilai signifikansi menunjukkan terdapat perbedaan yang berarti antara cluster 1, cluster 2 dan cluster 3 yang berhubungan dengan atribut-atribut pada penelitian.

2.4. Tahapan

Tahapan dalam pengerjaan makalah ini adalah : 1) deskripsi data dan penentuan metode *deadline* Jumat, 1 Oktober 2021. 2) proses analisis dan draft makalah minimal pendahuluan dan metode *deadline* Jumat, 8 Oktober 2021. 3) analisis lengkap dan *draft* makalah lengkap *deadline* Jumat, 15 Oktober 2021 dan presentasi dijadwalkan Senin, 18 Oktober 2021. 4) finalisasi makalah, Kamis 21 Oktober 2021.

3. Hasil dan Pembahasan

Dengan menggunakan bantuan software jupyter notebook python 3.0 diperoleh output sebagai berikut.

3.1. Tabel

Berikut ini adalah hasil analisisstatistik data saham dan analisis k-means.

Tabel 1. Analisis statistik Harga Saham 4 emiten

Statistic	WIKA	BISI	BSDE	NIKL
Length	132	132	132	132
Missing Count	0	0	0	0
Mean	1700.454700	1259.916667	1386.016879	1103.742424
Minimum	291.713135	490.000000	566.198853	50.000000
1st Quartile	972.377075	875.000000	1035.000000	188.500000
Median	1797.500000	1327.500000	1417.500000	335.000000
3rd Quartile	2372.502197	1665.000000	1735.000000	1335.000000
Maximum	3468.145020	2350.000000	2220.000000	5125.000000

Tabel 2. Analisis statistik Return Saham 4 emiten

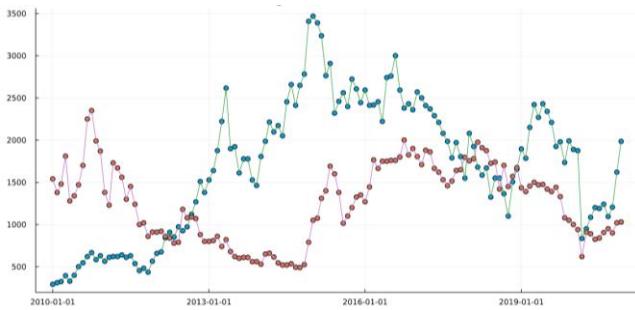
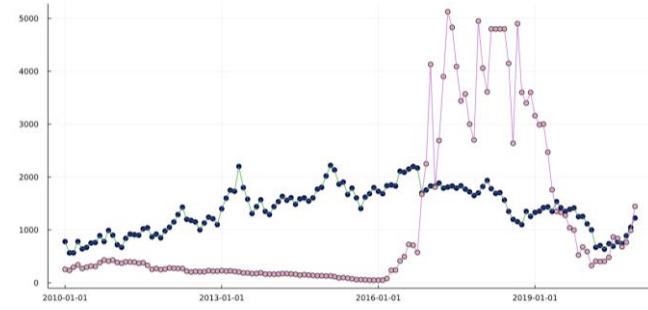
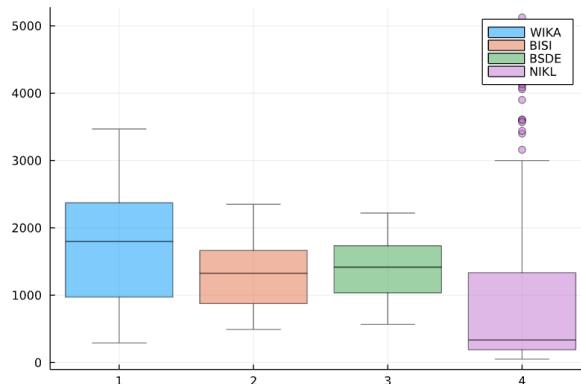
Statistic	WIKA	BISI	BSDE	NIKL
Length	132	132	132	132
Missing Count	0	0	0	0
Mean	0.023866	0.006780	0.009569	0.049343
Minimum	-0.554667	-0.340426	-0.330000	-0.561743
1st Quartile	-0.059424	-0.065380	-0.061317	-0.085671
Median	0.019334	0.000000	0.010807	-0.014760
3rd Quartile	0.104084	0.063016	0.076147	0.059635
Maximum	0.368182	0.504762	0.377049	1.913580

Tabel 3. Pergeseran titik centroid Harga Saham 4 emiten dalam 6 klaster

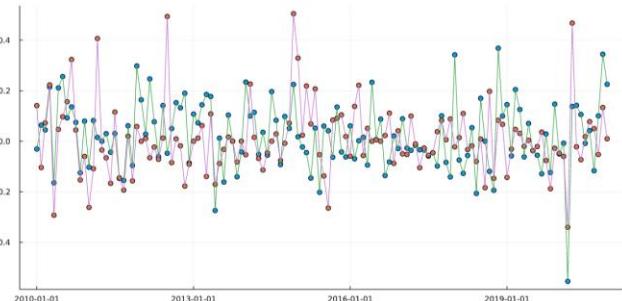
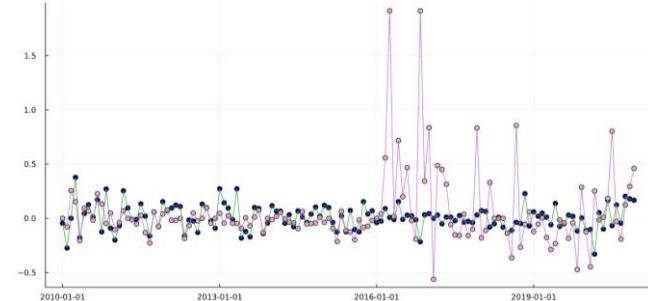
<i>Centroid</i>	<i>Klaster 0</i>	<i>Klaster 1</i>	<i>Klaster 2</i>	<i>Klaster 3</i>	<i>Klaster 4</i>	<i>Klaster 5</i>
<i>Centroid</i> pertama	233.559	2766.666	4875.625	767.307	3777.272	1520.714
<i>Centroid</i> kedua	233.559	3777.272	767.306	4875.625	2766.666	1520.714
<i>Centroid</i> ketiga	3777.272	233.559	1520.714	4875.625	2766.666	767.307
<i>Centroid</i> keempat	227.378	3777.272	1520.714	2766.666	729.933	4875.625
<i>Centroid</i> kelima	2766.666	233.559	4875.625	1520.714	3777.272	767.307
<i>Centroid</i> keenam	233.559	3777.272	1520.714	2766.666	4875.625	767.307

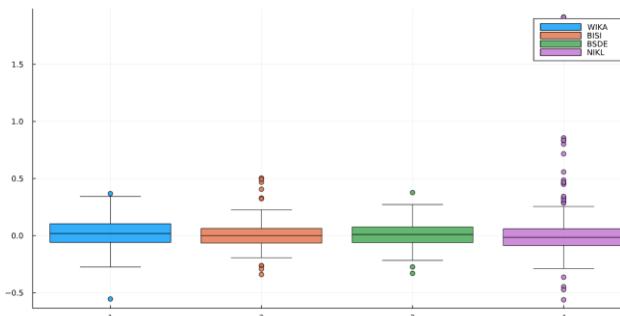
3.2. Gambar Analisis Data Harga Saham dan Return Saham

Gambar dari analisis data saham di bagi menjadi 2 yaitu gambar data harga saham dan data return saham. Warna orange menunjukkan data harga saham WIKA, sedangkan warna biru data harga saham BISI. Warna pink menunjukkan data harga saham BSDE, sedangkan warna hitam data harga saham NIKL.

**Gambar 1.** visualisasi data harga saham WIKA dan BISI**Gambar 2.** visualisasi data harga saham BSDE dan NIKL**Gambar 3.** boxplot data harga saham WIKA, BISI, BSDE dan NIKL

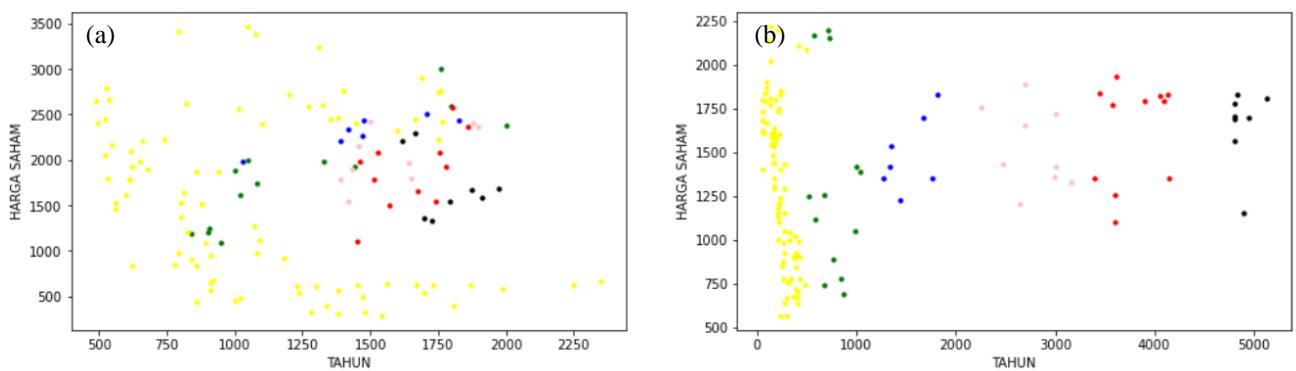
Sementara itu, gambar data data return saham. Warna orange menunjukkan data return saham WIKA, sedangkan warna biru data return saham BISI. Warna pink menunjukkan data return saham BSDE, sedangkan warna hitam data return saham NIKL.

**Gambar 4.** visualisasi data return saham WIKA dan BISI**Gambar 5.** visualisasi data return saham BSDE dan NIKL



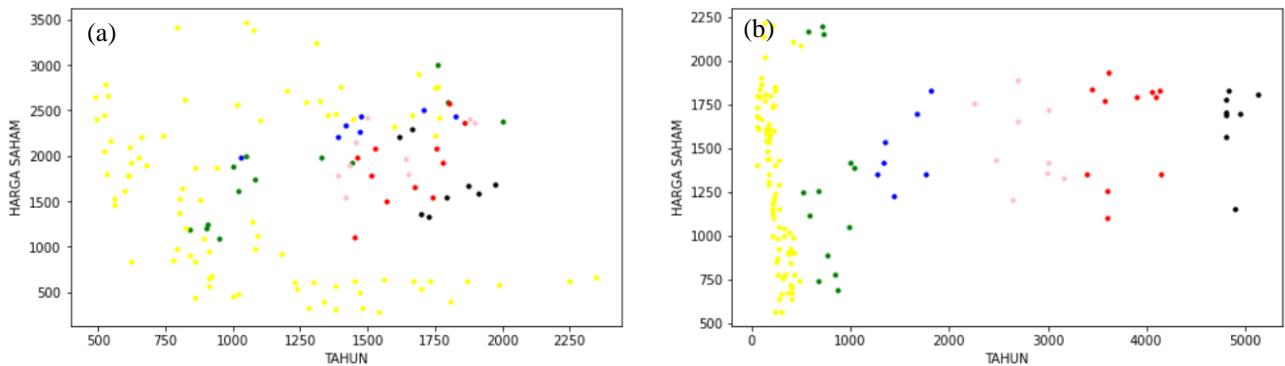
Gambar 6. boxplot data harga saham WIKA, BISI, BSDE dan NIKL

3.3 Visualisasi hasil clustering



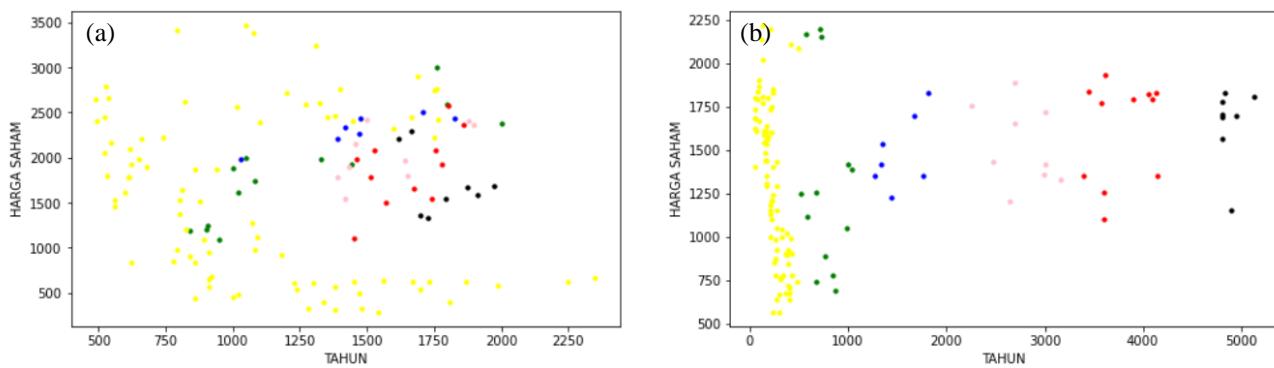
Gambar 7. Gambar: (a) visualisasi scatter prediksi pertama 6 klaster data harga saham WIKA dan BISI (b) visualisasi scatter prediksi kedua 6 klaster data harga saham BSDE dan NIKL

Warna hijau menunjukkan klaster ke-0, warna merah klaster ke-1, warna kuning klaster ke-2, warna hitam klaster ke-3, warna pink klaster ke-4, dan warna biru klaster ke-5. Titik centroid untuk klaster ke-0 adalah 233.5595238, klaster ke-1 adalah 2766.66666667, klaster ke-2 adalah 4875.625, klaster ke-3 adalah 767.30769231, klaster ke-4 adalah 3777.27272727, dan klaster ke-5 adalah 1520.71428571



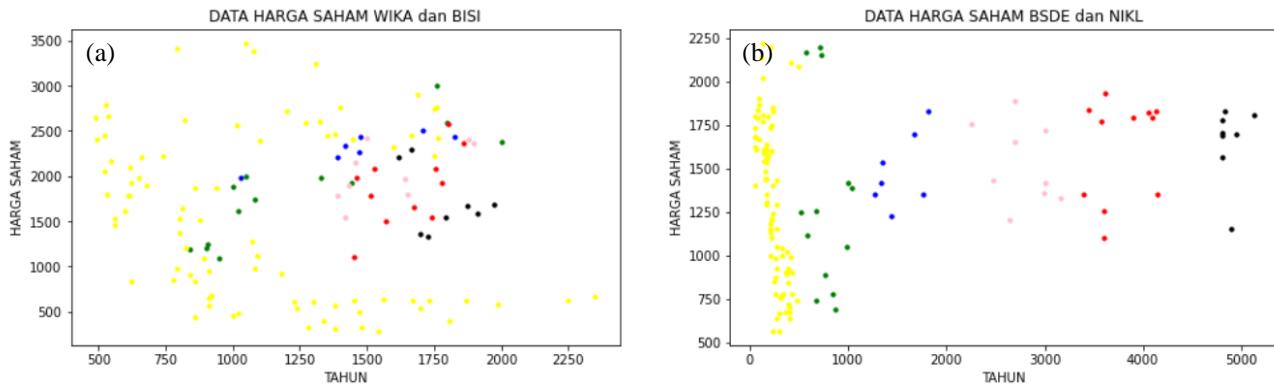
Gambar 8. Gambar: (a) visualisasi scatter prediksi kedua 6 klaster data harga saham WIKA dan BISI (b) visualisasi scatter prediksi kedua 6 klaster data harga saham BSDE dan NIKL

Titik centroid untuk klaster ke-0 adalah 233.55952381, klaster ke-1 adalah 3777.27272727, klaster ke-2 adalah 767.30769231, klaster ke-3 adalah 4875.625, klaster ke-4 adalah 2766.66666667, dan klaster ke-5 adalah 1520.71428571



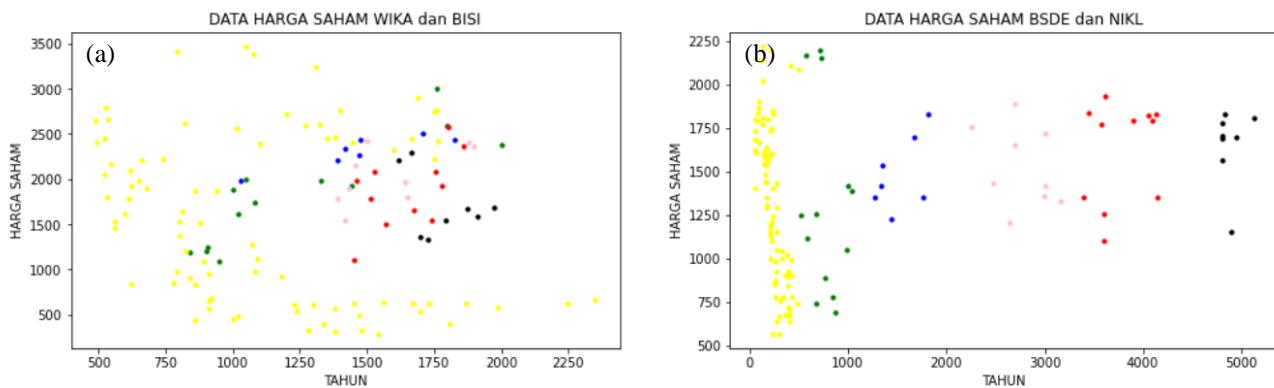
Gambar 9. Gambar: (a) *visualisasi scatter* prediksi ketiga 6 klaster data harga saham WIKA dan BISI (b) *visualisasi scatter* prediksi ketiga 6 klaster data harga saham BSDE dan NIKL

Sumbu X pada gambar (a) menunjukkan data harga saham BISI sedangkan sumbu Y menunjukkan data harga saham WIKA. Sedangkan sumbu X gambar (b) menunjukkan data harga saham NIKL dan sumbu Y menunjukkan data harga saham BSDE. Titik centroid untuk klaster ke-0 adalah 3777.27272727, klaster ke-1 adalah 233.55952381, klaster ke-2 adalah 1520.71428571, klaster ke-3 adalah 4875.625, klaster ke-4 adalah 2766.66666667, dan klaster ke-5 adalah 767.30769231



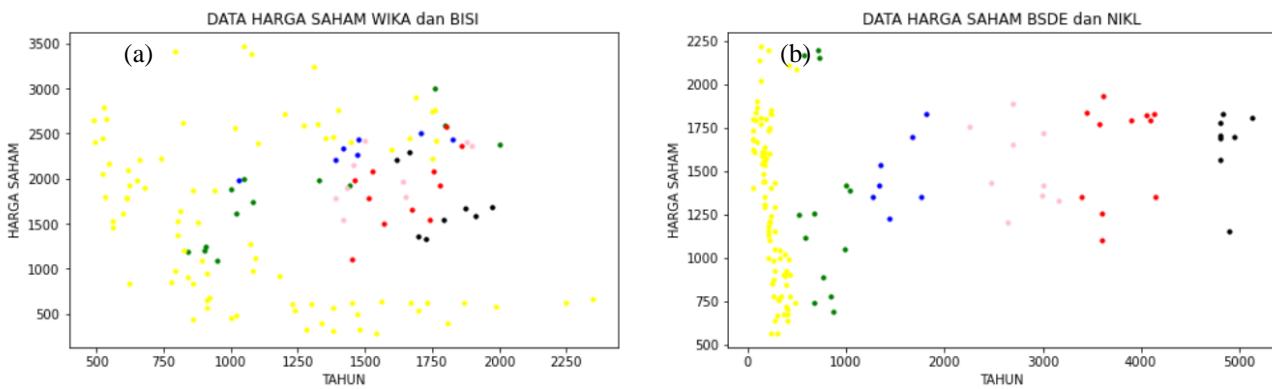
Gambar 10. Gambar: (a) *visualisasi scatter* prediksi keempat 6 klaster data harga saham WIKA dan BISI (b) *visualisasi scatter* prediksi keempat 6 klaster data harga saham BSDE dan NIKL

Titik centroid untuk klaster ke-0 adalah 227.37804878, klaster ke-1 adalah 3777.27272727, klaster ke-2 adalah 1520.71428571, klaster ke-3 adalah 2766.66666667, klaster ke-4 adalah 729.93333333, and klaster ke-5 adalah 4875.625



Gambar 11. Gambar: (a) *visualisasi scatter* prediksi kelima 6 klaster data harga saham WIKA dan BISI (b) *visualisasi scatter* prediksi keempat 6 klaster data harga saham BSDE dan NIKL

Titik centroid untuk klaster ke-0 adalah 2766.66666667, klaster ke-1 adalah 233.55952381, klaster ke-2 adalah 4875.625, klaster ke-3 adalah 1520.71428571, klaster ke-4 adalah 3777.27272727, and klaster ke-5 adalah 767.30769231



Gambar 12. Gambar: (a) *visualisasi scatter* prediksi keenam 6 klaster data harga saham WIKA dan BISI (b) *visualisasi scatter* prediksi keempat 6 klaster data harga saham BSDE dan NIKL

Titik centroid untuk klaster ke-0 adalah 233.55952381, klaster ke-1 adalah 3777.27272727, klaster ke-2 adalah 2766.66666667, klaster ke-3 adalah 1520.71428571, klaster ke-4 adalah 4875.625, dan klaster ke-5 adalah 767.30769231.

4. Simpulan

Metode K-means dapat mengklasifikasi data harga saham ke dalam 6 klaster yang dipengaruhi oleh titik centroid yang dihasilkan. Metode ini bertujuan untuk mempartisi data yang ada ke dalam kelompok sehingga data berkarakteristik sama dimasukan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokan kedalam kelompok yang lain. Nilai signifikansi menunjukkan terdapat perbedaan yang berarti antara klaster 0, klaster 1, klaster 2, klaster 3, klaster 4 dan klaster 5 yang berhubungan dengan atribut-atribut pada penelitian.

5. Ucapan Terima Kasih

Penulis menyampaikan ucapan terima kasih kepada Tuhan Yang Maha Esa atas berkat dan bimbingan-Nya sehingga penulisan makalah ini dapat diselesaikan. Ucapan terimakasih juga kepada teman-teman angkatan 58 khususnya kepada Ade Irawan dan Dilah Rismawati atas bantuan informasi dan diskusi. Ucapan terima kasih juga saya sampaikan kepada M.K. Najib atas masukan terhadap pemilihan metode yang cocok dalam menganalisis data saham.

6. Daftar Pustaka

- Joshi, Anshul. 2016. Julia for Data Science.Birmingham: Packt Publishing Ltd.
- Hair, J. F, Anderson, R. E, Tanham, R. L, and Black, W. C., 1998. Multivariate data analysis. Fifth Edition.Prentice Hall International, Inc. Upper Saddle River,New Jersey.
- Sibarani , Rama. 2018. Algorithma K-Means Clustering Strategi Pemasaran Penerimaan Mahasiswa Baru Universitas Sty Negara Indonesia. Jurnal Algoritma, Logika dan Komputasi. Vol.I (No. 2) : 44 - 50.
- Suryana, Nana.2011. Penggunaan Metode Statistik K-Means Clustering Pada Analisis Peruntukan Lahan Usaha Tambang Berbasis Sistem Informatika Geografi. Jurnal Teknologi Mineral dan Batubara Volume 7, Nomor 1, Januari 2011 : 42 – 53
- Putri Utami, Mailia. 2019. Pengaruh Teknik Clustering Harga Saham Dalam Manajemen Portofolio.e-Proceeding of Engineering : Vol.6, No.1 April 2019.Page 2491
- Rama Putra, Andhika, dkk. 2019. Pembentukan Portofolio Saham Melalui Proses Clustering Kurva Harga Saham Hasil Spline Kuadratik. e-Proceeding of Engineering : Vol.6, No.2. Page 9867

Pengaplikasian Metode *Autoregressive* Dalam Memprediksi Data Saham

Ade Irawan

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: ade150499irawan@apps.ipb.ac.id

Abstrak.

Saham merupakan instrument investasi yang banyak dipilih oleh para investor karena saham mampu memberikan tingkat keuntungan yang sangat tinggi, bahkan investasi saham dapat memberikan hasil yang maksimal apabila dikontrol dengan manajemen risiko yang tepat. Data harga saham merupakan data runtun waktu sehingga dalam melakukan analisis data saham dapat menggunakan metode runtun waktu klasik. Namun, jika menggunakan metode runtun waktu klasik akan banyak asumsi-asumsi yang harus terpenuhi. Berdasarkan hal tersebut, dalam melakukan analisis data saham diperlukan metode alternatif seperti metode *autoregressive*. Hal tersebut dikarenakan metode *autoregressive* tidak mengharuskan adanya asumsi khusus sehingga mudah digunakan dalam menganalisis data harga saham. Metode *autoregressive* adalah salah satu model regresi yang menyatakan bahwa data pada periode sekarang dipengaruhi oleh data pada periode sebelumnya. Penulisan makalah ini bertujuan mengetahui keakuratan metode *autoregressive* dalam melakukan prediksi data harga saham bulanan dari WIKA, BISI, BSDE, dan NIKL. Data yang digunakan dalam penulisan makalah ini merupakan data historis harga saham bulanan sejak Januari 2010 sampai Desember 2019. Dari hasil analisis dengan menggunakan metode *autoregressive* menunjukkan bahwa data harga saham sejak Januari 2010 sampai Desember 2019 dapat digunakan untuk memprediksi harga saham di masa mendatang. Berdasarkan hasil analisis yang telah dilakukan dengan menggunakan *software Julia 1.6.2* dapat disimpulkan bahwa data hasil prediksi cenderung mengikuti data historis yang digunakan atau mengalami kenaikan dan penurunan yang tidak mempunyai pola tertentu.

Kata Kunci: Metode Autoregressive, Data Saham.

1. Pendahuluan

Saham adalah salah satu jenis surat berharga yang digunakan sebagai tanda penyerta modal seseorang atau badan usaha dalam suatu perusahaan (Satriawan, 2017). Sebelum melakukan investasi, penting bagi investor untuk mengetahui seberapa besar risiko dan *return* dari saham tersebut. Investasi merupakan suatu kegiatan jual beli pada zaman modern yang paling diminati di era globalisasi. Sasaran investasi yang paling diminati oleh para investor yaitu investasi saham karena dapat memberikan keuntungan yang sangat besar di masa mendatang (Rahmi & Mahmudy. 2016). Untuk mendapatkan keuntungan yang besar, investor harus jeli dalam mengolah sahamnya karena jika tidak investor akan mendapatkan risiko kerugian yang besar. Dalam meminimalisir terjadinya risiko kerugian, investor perlu melakukan prediksi data saham misalkan seperti harga saham. Menurut Rahmi & Mahmudy (2016) menyatakan prediksi harga saham yang tepat dan benar dengan kondisi menguntungkan untuk investor, dapat dilakukan melalui pendekatan analisis teknikal. Berdasarkan data saham yang digunakan dalam kasus ini, penulis menggunakan model autoregresif dengan data saham masa lalu untuk memprediksi data saham di masa mendatang.

Model *autoregressive* adalah salah satu model regresi yang menyatakan bahwa data pada periode sekarang dipengaruhi oleh data pada periode sebelumnya (Lilipaly, dkk. 2014). Model autoregresif merupakan suatu model yang menjelaskan gerakan variabel melalui variabel di masa lalu. Selain menggunakan model autoregresif, prediksi harga saham juga dapat dilakukan dengan menggunakan pendekatan *artificial intelligence* (Rahmi & Mahmudy. 2016). Pendekatan *artificial intelligence* yang dimaksud adalah pendekatan yang diterapkan dengan model regresi.

Penelitian data saham menggunakan metode autoregresif yang pernah dilakukan sebelumnya adalah “prediksi harga saham PT. BRI, Tbk. Menggunakan metode ARIMA (*Autoregressive Integrated Moving Average*) (Lilipaly, 2016). Dalam penelitian tersebut, peneliti memfokuskan kepada pembuatan model ARIMA dan memprediksi harga saham PT. BRI, Tbk. Bulan November 2014. Selanjutnya penelitian yang dilakukan oleh Prabawa, Nasri, & Sulistiyo (2015) yang melakukan “prediksi harga saham dengan menggunakan metode *autoregressive* dan algoritma kelelawar”. Penelitian yang dilakukan oleh Prabawa, Nasri, & Sulistiyo (2015) berfokus pada penyelesaian masalah prediksi harga saham dengan menggunakan metode *Autoregressive* dan algoritma kelelawar. Kemudian penelitian yang dilakukan oleh Rusyida & Pratama (2020) yang melakukan “prediksi harga saham Garuda Indonesia di tengah pandemi *Covid-19* menggunakan metode ARIMA”. Penelitian tersebut berfokus dalam mengetahui keakuratan metode ARIMA dalam melakukan peramalan atau prediksi harga saham harian PT. Garuda Indonesia, Tbk di tengah pandemi *Covid-19*.

Berdasarkan penelitian-penelitian tersebut, permasalahan-permasalahan dalam memprediksi harga saham dapat dipecahkan menggunakan metode *autoregressive*. Dari beberapa data saham yang di lihat, pada kenyataannya pergerakan data saham tidak membentuk pola secara linear. Berdasarkan hal tersebut, penggunaan metode

autoregressive dalam memprediksi harga saham sangat mendukung karena metode *autoregressive* merupakan suatu model yang dapat mengikuti pergerakan data *historis*. Dari uraian tersebut, penulis bertujuan untuk melakukan prediksi data saham khususnya pada data harga saham dengan menggunakan metode *autoregressive*.

2. Data dan Metode

Data yang digunakan dalam penulisan makalah ini adalah data saham masa lalu (data tahun 2010-2020). Data tersebut berupa data harga saham dan data return saham. Dalam makalah ini, data yang akan dianalisis yaitu data harga saham. Dalam melakukannya analisis data harga saham, penulis menggunakan metode *autoregressive* untuk memprediksi terhadap data harga saham. Menurut Pawestri, dkk (27: 2019) menyatakan bahwa metode *autoregressive* adalah suatu bentuk model regresi yang bukan menghubungkan variabel tak bebas, melainkan model regresi yang dapat menghubungkan nilai-nilai sebelumnya pada selang waktu yang bermacam-macam.

2.1. Alat dan Bahan

Adapun alat dan bahan yang digunakan dalam melakukan prediksi data harga saham adalah sebagai berikut.

1. Personal computer (PC)

Spesifikasi komputer yang digunakan untuk menganalisis prediksi data harga saham dapat di lihat pada Tabel 1 berikut.

Tabel 1. Spesifikasi computer

No.	System	
1	Operating System	Windows 10 Pro 64-bit (10.0, Build 19042)
2	Computer Name	DESKTOP PEIUOEK
3	System Manufacturer	Dell Inc.
4	System Model	Inspiron 14-3467
5	BIOS	2.9.0
6	Processor	Intel® Core™ i3-7020U CPU @ 2.30GHz (4CPUs), -2.3GHz
7	Diretx Version	DirectX 12
8	Memory	4096MB RAM

2. Software Julia 1.6.2

Software yang digunakan untuk menganalisis prediksi data harga saham dalam makalah ini adalah *Software Julia 1.6.2*. Dalam penggunaan *software* tersebut, penulis menggunakan beberapa *package* khusus untuk melakukan analisis data. Adapun *package* yang digunakan dapat di lihat pada Tabel 2.2 berikut.

Tabel 2. Package Julia

No.	Package	Function
1	Plots	Menunjukkan tampilan gambar sebaran data
2	DataFrame	Mendeskripsikan suatu data dari data keseluruhan
3	XLSX	Mendeskripsikan data dari Microsoft Excel
4	RDatasets	Mendeskripsikan data
5	StatsPlots	Menunjukkan bentuk gambar violin
6	StatsBase	Deskripsi statistic data saham baik harga maupun return
7	VMLS	Memprediksi data harga saham

2.2. Data

Data adalah *things known or assumed* (Situmorang *et al.* 2010). Data yang digunakan adalah data saham dari daftar harga dan *return* dari empat saham di Indonesia (data dari Yahoo! Finance). Menurut Pratama & Rusyida (2020) data tersebut merupakan jenis data sekunder. Khusus dalam makalah ini, data yang digunakan merupakan data harga saham dari empat saham di Indonesia pada tahun 2010-2020. Data tersebut merupakan data *time series* (kurun waktu) karena data tersebut diambil selama kurun waktu sebelas tahun dan data tersebut merupakan data kuantitatif.

2.3 Penggunaan Violin Plot untuk Data Harga Saham

A violin plot is essentially a smoothed histogram rotated by 90° that provides a density estimation of these data (Kenny & Schoen, 2021). “A violin plot that shows the distribution of the predicted score, by outcome for the training (Mendez, *et.al.* 2020)”. Dalam makalah ini, *violin plot* akan digunakan untuk membuat histogram data halus dari data harga dan *return* saham.

2.4 Metode Autoregressive

Metode *Autoregressive* pertama kali diperkenalkan oleh Yule pada Tahun 1926 dan dikembangkan oleh Walker pada Tahun 1931, model ini memiliki asumsi bahwa data periode sekarang dipengaruhi oleh data pada periode sebelumnya (Hatidja, 2011). Metode *autoregressive* adalah salah satu model regresi yang menyatakan bahwa data pada periode sekarang dipengaruhi oleh data pada periode sebelumnya (Lilipaly, dkk. 2014). Menurut Makridakis (dalam Pawestri, dkk. 2019) suatu metode *autoregressive* dapat menyatakan suatu prediksi sebagai fungsi nilai-nilai sebelumnya dari *time series* tertentu. Selain memprediksi harga saham, metode *autoregressive* dapat memodelkan harga saham (Prabawa, dkk. 2015). Metode *autoregressive* dengan order p dinotasikan dengan AR(p) dengan bentuk umum berikut (Lusikoy, dkk. 2019).

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

dengan,

y_t	: Nilai variabel pada waktu ke- t , $t = 1, 2, \dots, n$.
$y_{t-1}, y_{t-2}, \dots, y_{t-p}$: Nilai masa lalu dari <i>time series</i> yang bersangkutan pada waktu ke $t - 1, t - 2, \dots, t - p$
ϕ_i	: Parameter AR tingkat i , $i = 1, 2, 3, \dots, p$
ε_t	: Nilai <i>error</i> pada waktu ke- t
p	: Order AR

Analisis data dilakukan dengan menggunakan metode *autoregressive* dengan bantuan *software Julia 1.6.2*. Adapun langkah-langkah dalam menganalisis data saham tersebut secara berturut-turut adalah (Justica, dkk. 2021).

1. Pengambilan data;
2. Penentuan karakteristik data;
3. Plot data
4. Melakukan prediksi

3. Hasil dan Pembahasan

Saham merupakan salah satu instrumen investasi yang memiliki tingkat risiko yang lebih tinggi dibandingkan dengan instrumen investasi lain, tetapi investasi saham memberi hasil yang maksimal apabila dikontrol dengan manajemen risiko yang tepat (Hayati, 2016). Pada bab ini, penulis akan membahas mengenai pemrediksian jenis data harga saham yaitu WIKA, BISI, BSDE, dan NIKL dengan menggunakan metode *autoregressive* untuk pemrediksian data harga saham di masa mendatang.

3.1 Pengambilan Data

Data saham yang digunakan adalah data harga saham jenis WIKA, BISI, BSDE, dan NIKL Tahun 2010-2020. Data tersebut merupakan data *time series* dan digunakan untuk melakukan prediksi harga saham di masa mendatang.

3.2 Karakteristik Data Harga Saham

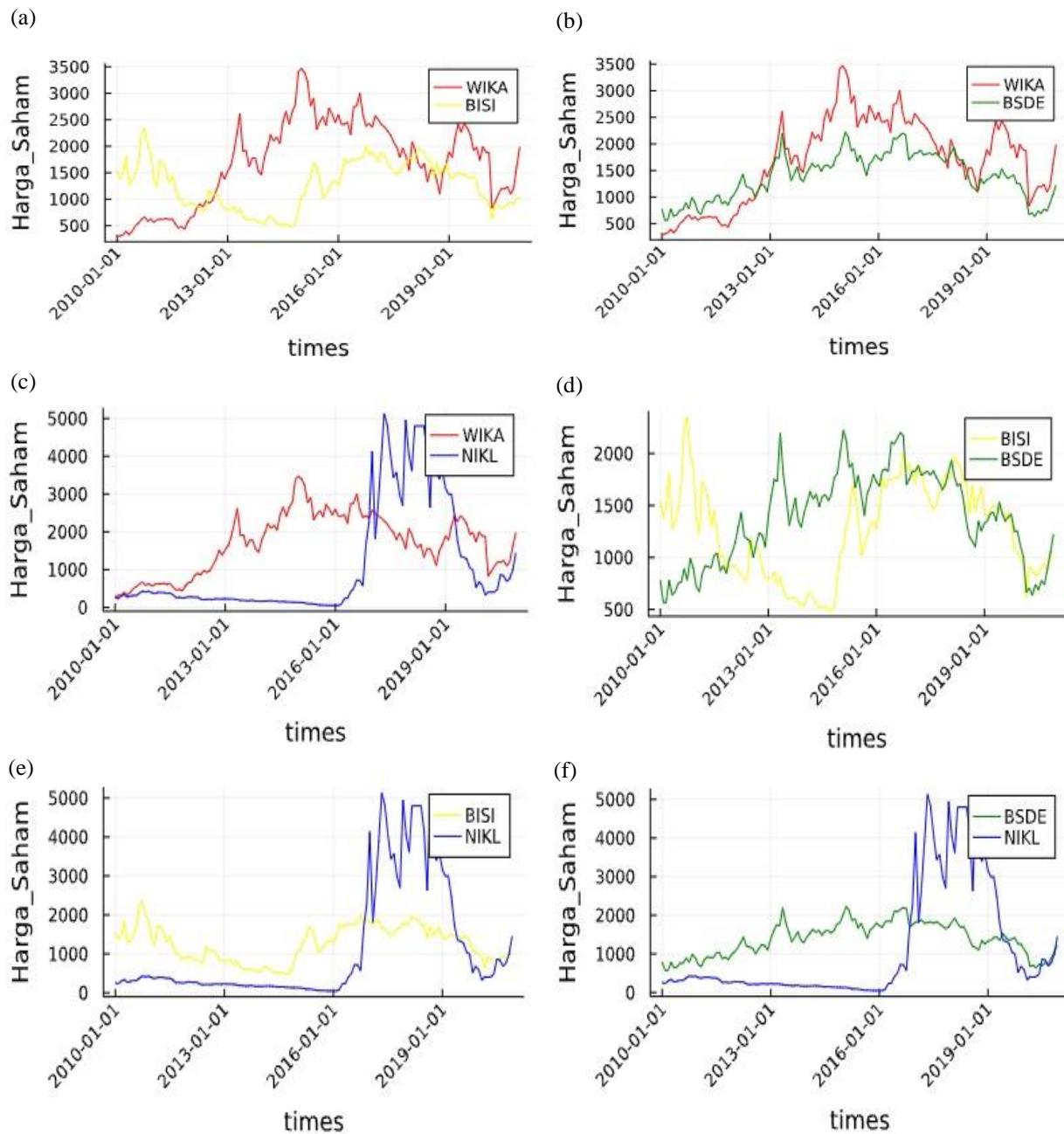
Secara umum, gambaran data harga saham dapat diketahui melalui statistik deskriptif. Adapun karakteristik untuk menggambarkan informasi dari data harga saham WIKA, BISI, BSDE, dan NIKL dapat di lihat pada Tabel 3 berikut.

Tabel 3. Karakteristik data harga saham

No.	Variabel	Mean	Minimum	Maximum	Median
1	WIKA	1700.454700	291.713135	3468.145020	1797.500000
2	BISI	1259.916667	490.000000	2350.000000	1327.500000
3	BSDE	1386.016879	566.198853	2220.000000	1417.500000
4	NIKL	1103.742424	50.000000	5125.000000	335.000000

3.3 Plot Data

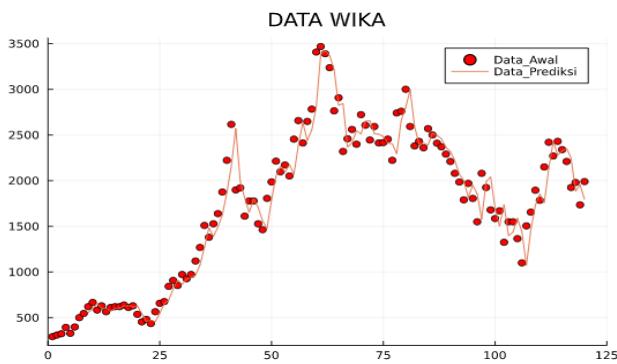
Pergerakan data harga saham jenis WIKA, BISI, BSDE, dan NIKL pada Bulan Januari Tahun 2010 sampai dengan Bulan Desember Tahun 2020 dapat diketahui melalui bentuk *time series* plot. Plot *time series* pada gambar di bawah ini akan menunjukkan gerakan yang tidak tertentu atau pada data harga saham tersebut tidak menunjukkan pola tertentu. Hal ini dapat disebabkan karena beberapa faktor seperti kondisi harga saham setiap bulannya. Selanjutnya, setelah data di plot menggunakan *software Julia* menunjukkan sebuah grafik yang naik turun seperti pada Gambar 1 berikut.



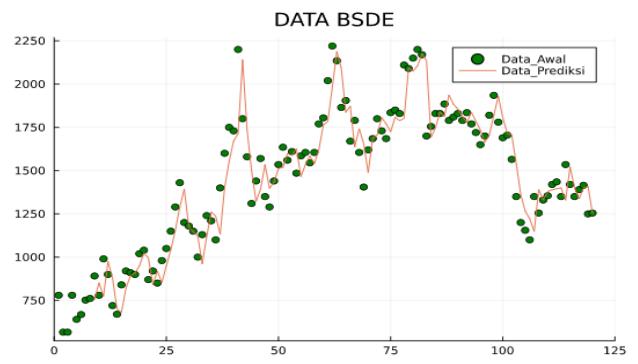
Gambar 1. Time series plots (a) WIKA & BISI, (b) WIKA & BSDE, (c) WIKA & NIKL, (d) BISI & BSDE, (e) BISI & NIKL, dan (f) BSDE & NIKL

3.4 Prediksi

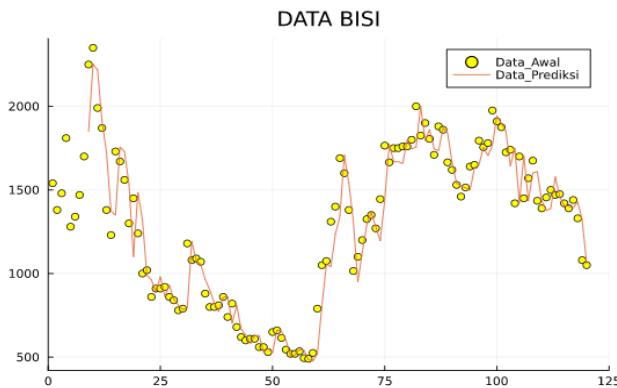
Model regresi yang terbaik dapat digunakan untuk melakukan suatu prediksi untuk melihat suatu data di masa mendatang (Puspitasari, dkk. 2012). Dalam kasus ini, data harga saham yang sudah diketahui digunakan untuk melakukan prediksi dari data harga saham. Selanjutnya, setelah dilakukan analisis prediksi data harga saham diperoleh suatu grafik yang mengalami kenaikan dan penurunan yang hampir mendekati data awal yang digunakan, tetapi tidak mempunyai pola tertentu. Hal tersebut dikarenakan data harga saham yang digunakan sulit untuk diprediksi (Kartikasari, 2020). Adapun grafik prediksi (*forecasting*) harga saham dapat di lihat pada Gambar berikut.



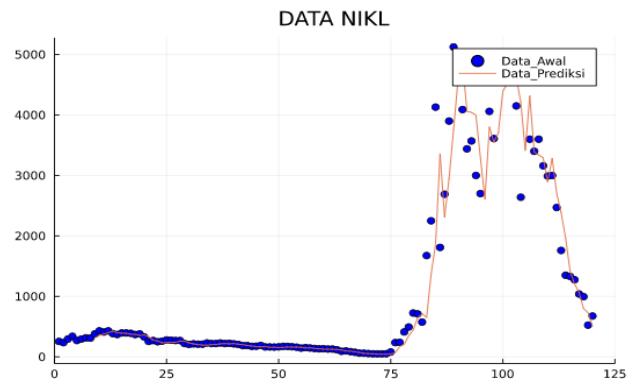
Gambar 2. Prediksi data harga saham WIKA



Gambar 3. Prediksi data harga saham BSDE



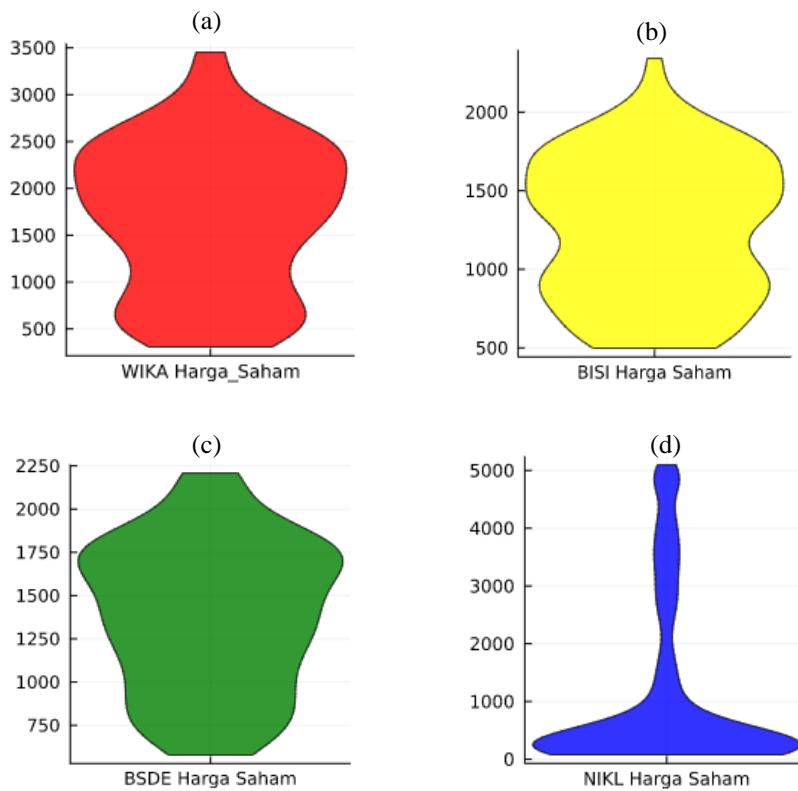
Gambar 4. Prediksi data harga saham BISI



Gambar 5. Prediksi data harga saham NIKL

3.5 Violin Plot

Berdasarkan hasil *violin plot* data harga saham menggunakan *software Julia 1.6.2* berbantuan *package StatsPlots* diperoleh hasil berikut.



Gambar 5. *Violin plots* data harga saham (a) WIKA (b) BISI, (c) BSDE, dan (d) NIKL.

4. Simpulan

Berdasarkan hasil analisis data harga saham tersebut, dapat disimpulkan bahwa hasil perhitungan karakteristik dengan menggunakan *software Julia 1.6.2* dapat diperoleh bahwa data harga saham WIKA mempunyai rata-rata (*mean*) paling tinggi yaitu sebesar 1700.454700, sedangkan yang mempunyai rata-rata (*mean*) terendah adalah data harga saham NIKL dengan rata-rata (*mean*) sebesar 1103.742424. Pergerakan data awal harga saham dengan data hasil prediksi cenderung memiliki kenaikan dan penurunan yang hampir sama. Secara keseluruhan, data harga saham tertinggi terjadi pada data harga saham NIKL yang mencapai lebih dari 5.000, sedangkan yang mempunyai harga saham terendah terjadi pada data harga saham NIKL juga yang hampir mencapai titik 0.

5. Ucapan terima kasih

Melalui kesempatan ini, penulis menyampaikan banyak terima kasih yang sebesar-besarnya kepada Saudari Anggun Praptaningsih, Refi Revina, dan Trianty Putri Blante yang sudah meluangkan sedikit waktunya untuk melakukan diskusi dalam pembahasan cara mengaplikasikan metode *autoregressive* ke *software Julia 1.6.2* sehingga pengaplikasian metode tersebut dapat terselesaikan pada waktu yang tepat.

6. Daftar Pustaka

- Hatidja D. 2011. Penerapan Model ARIMA Untuk Memprediksi Harga Saham PT. TELKOM Tbk. *Jurnal Ilmiah Sains*. 11(1):116–123.
doi: <https://doi.org/10.35799/jis.11.1.2011.53>
- Hayati FN. 2016. Peramalan Harga Saham Jakarta Islamic Index Menggunakan Metode Vector Autoregressive [Skripsi]. Institut Teknologi Sepuluh November.
- Justica A, Adzkiya A, Fadhilah A, Pravitasari AA, Indrayatna F. 2021. Prediksi Harga Saham UNVR Menggunakan Autoregressive Integrated Moving Average. *Seminar Nasional Statistika Online (SNSO 2021)*.
- Kartikasari P. 2020. Prediksi Harga Saham PT. Bank Negara Indonesia Dengan Menggunakan Model Autoregressive Fractional Integrated Moving Average (ARFIMA). *Jurnal Unimus*. 8(1):1-7.
- Kenny M, Schoen I. 2021. Violin SuperPlots: Visualizing Replicate Heterogeneity in Large Data Sets. *School of Pharmacy and Biomolecular Sciences*. 32:1333-1334.
- Lilipaly GS, Hatidja D, Kekenus JS. 2014. Prediksi Harga Saham PT. BRI, Tbk. Menggunakan Metode ARIMA (Autoregressive Integrated Moving Average). *Jurnal Ilmiah Sains*. 14(2):60-67.
doi: <https://doi.org/10.35799/jis.14.2.2014.5927>
- Lusikooy J, Nainggolan N, Titaley J. 2017. Prediksi Harga Tutup Saham PT. Garuda Indonesia, Tbk. Menggunakan Metode ARIMA. *Jurnal MIPA UNSRAT Online*. 6(1):74-77.
doi: <https://doi.org/10.35799/jm.6.1.2017.16174>
- Mendez KM, Broadhurst DI, Reinke SN. 2020. Migrating from partial least squares discriminant analysis to artificial neural networks: a comparison of functionally equivalent visualisation and feature contribution tools using jupyter notebooks. *Metabolomics*. 16(17):1-15.
doi: <https://doi.org/10.1007/s11306-020-1640-0>
- Pawestri V, Setiawan A, Linawati L. 2019. Pemodelan Data Penjualan Mobil Menggunakan Model Autoregressive Moving Average Berdasarkan Metode Bayesian. *Jurnal Sains dan Edukasi Sains*. 2(1):26-35.
doi: <https://doi.org/10.24246/juses.v2i1p26-35>
- Prabawa B, Nasri J, Sulistiyo MD. 2015. Prediksi Harga Saham Dengan Menggunakan Metode Autoregressive dan Algoritma Kelelawar. *e-Proceding of Engineering*. 2(1):1696-1704.
- Pratama VY, Rusyida WY. 2020. Analisis Prediksi Harga Saham Sinar Mas Group. *Jurnal Pasar Modal dan Bisnis*. 2(2):203-216.
doi: <https://doi.org/10.37194/jpmb.v2i2.52>
- Puspitasari I, Suparti, Wilandari Y. 2012. Analisis Indeks Harga Saham Gabungan (IHSG) Dengan Menggunakan Model Regresi Kernel. *Jurnal Gaussian*.1(1):93-102.
doi: <https://doi.org/10.14710/j.gauss.v1i1.577>
- Rahmi A, Mahmudy WF. 2016. Pembentukan Model Regresi Harga Saham Menggunakan Algoritma Genetika. *Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016)*.
- Rusyida WY, Pratama VY. 2020. Prediksi Harga Saham Garuda Indonesia di Tengah Pandemi Covid-Menggunakan Metode ARIMA. *Journal of Mathematics and Mathematics Education*.2(1):78-81.
doi: [10.21580/square.2020.2.1.5626](https://doi.org/10.21580/square.2020.2.1.5626)
- Satriawan IW. 2017. Prediksi Volatilitas Saham Perusahaan Pertambangan Batu Bara Dengan Metode Artificial Neural Networks-Generalized Autoregressive Conditional Heteroscedasticity. *e-Proceding of Engineering*. 4(3):5184-5195.
- Situmorang [et. al.]. 2010. *Analisis Data: untuk riset manajemen dan bisnis*. Medan: USU Press 2010.
Yahoo.Finance.com.

Estimasi Trend dan Seasonal Component Data Curah Hujan di Kalimantan

Refi Revina

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: refirevina@apps.ipb.ac.id

Abstrak.

Dalam makalah ini, akan dikaji mengenai curah hujan dari beberapa kota di Kalimantan. Hujan sendiri merupakan salah satu penilaian terhadap kategori cuaca. Cuaca adalah keadaan udara di atmosfer pada waktu dan tempat tertentu yang memiliki sifat berubah-ubah atau tidak tetap. Penilaian terhadap kategori cuaca sendiri umumnya dinyatakan dengan memperhatikan kondisi berupa hujan, suhu udara, jumlah tutupan awan, penguapan kelembaban, dan kecepatan angin di suatu tempat dari pada suatu satuan waktu. Cuaca dikatakan menjadi salah satu hal yang penting dalam kehidupan dan di berbagai sektor, seperti pertanian, industri, dan lain-lain. Begitu pula dengan hujan, sehingga fenomena tersebut sangat menarik untuk dikaji lebih lanjut. Data yang tersedia merupakan data curah hujan dalam periode Januari 2001-Desember 2019 dan dinyatakan menjadi curah hujan per bulan yang menjadikan data ini sebagai data dengan pola *time series* yang memuat data 15 Kota/Kab di Kalimantan. Selanjutnya, akan ditelaah mengenai estimasi dari *Trend* dan *Seasonal Component* dari data curah hujan dari kota yang telah dipilih.

Kata Kunci: curah hujan, *trend*, *seasonal*.

1. Pendahuluan

Hujan merupakan bentuk presipitasi atau proses terjadinya pengendapan air dari atmosfer ke permukaan bumi dalam bentuk cair (hujan) dan padat (salju) (Harijono *et al.* 2010). Intensitas hujan dipengaruhi oleh banyak faktor seperti letak geografis, topografi, arah angin dan lain-lain. Sebagian besar wilayah Indonesia termasuk daerah dengan tipe hujan monsun dan merupakan daerah konvektif paling aktif di dunia (Ramage 1971; Webster 1987; Tjasyono *et al.* 2006). Hal ini dapat dilihat dari tingginya aktivitas awan konvektif di daerah tersebut dan pola hujan rata-ratanya yang mengikuti siklus aktivitas monsun, terutama monsun Asia dan Australia. Variabilitas curah hujan di beberapa provinsi di Indonesia lebih dominan dikendalikan oleh *El Nino-Southern Oscillation* (ENSO) daripada *Indian Dipole Mode* (IOD) yang sebagian besar mendorong variabilitas curah hujan musiman dan antar-tahunan serta fenomena alam ekstrem di Indonesia lainnya (Prasetya R 2011; Satya Wardhana H *et al.* 2016).

Interaksi monsun Asia-Australia menyebabkan sebagian besar musim hujan di Indonesia dapat dikategorikan dengan jelas, yaitu musim hujan pada bulan Desember-Januari-Februari (DJF) dan musim kemarau pada bulan Juni-Juli-Agustus (JJA). Namun, variasi curah hujan di Indonesia tidak sama untuk setiap daerah. Untuk beberapa daerah yang dekat dengan garis khatulistiwa yang melewati beberapa area di Sumatera, Kalimantan dan Sulawesi maka komponen siklus hujan 6 bulan cukup kuat yang ditandai dengan adanya dua nilai maksimal di sekitar periode ekuinoks, yaitu pada bulan Maret-April dan Oktober-November (Chang *et al.* 2005). Sederhananya, dapat dipahami bahwa pada Kalimantan Barat terjadi dua kali musim hujan per tahunnya.

Data yang digunakan pada makalah ini merupakan data curah hujan dari kota Samarinda, Tanjung Selor, Palangkaraya, Banjarmasin, dan Sintang. Data tersebut merupakan data yang diambil selama sepuluh tahun dari periode Januari 2001-Desember 2019 atau juga dapat disebut sebagai data *time series* dengan pola data musiman. Data berkala (*time series*) adalah data yang disusun berdasarkan urutan waktu atau data yang dikumpulkan dari waktu ke waktu. Waktu yang digunakan dapat berupa minggu, bulan, tahun dan sebagainya. *Time Series* merupakan serangkaian data pengamatan yang berasal dari satu sumber tetap dan terjadi berdasarkan indeks waktu *t* secara beruntun dengan interval waktu yang tetap (Cryer, 1986).

Trend adalah keadaan data yang naik atau turun sering waktu ke waktu. *Trend* dapat dibagi dua yaitu *trend* yang terpengaruh oleh pola data musiman dan *trend* yang tidak terpengaruh oleh pola data musiman. Dalam makalah ini, dari data curah hujan beberapa kota di Kalimantan dan akan dilakukan estimasi *Trend* dan *Seasonal Component*, mengamati apakah *trend* naik atau turun, lalu akan dilakukan perbandingan *trend* tanpa pengaruh musim dan *trend* dengan pengaruh musim dengan mengamati grafik dan dengan mempertimbangkan nilai *Root Mean Square Error* (RMSE) dari selisih *trend* dan data yang ada. Pada penelitian sebelumnya dengan data yang sama telah dilakukan analisis pola curah hujan di Kalimantan menggunakan *fast fourier transform* (FFT) dan *empirical orthogonal functions* (EOF) (S Nurdiani *et al.* 2021).

2. Data dan Metode

2.1. Alat dan Bahan

Berikut ini merupakan alat dan bahan yang menunjang proses penggerjaan makalah ini.

1. Personal Computer dengan spesifikasi sebagai berikut :

Nama	ASUS
System Model	TX201LA
Processor	Intel(R) Core(TM) i7-4500U CPU @ 1800GHz (4CPUs), ~2.4GHZ
Display	11.6" IPS Full HD Glossy Touch Screen
Sistem Operasi	Windows 8.1 Single Language 64-bit (6.3, Build 9600) + Android 4.2
RAM	4.00 GB
Konektivitas	Wi-Fi/Camera/Card Reader/USB/Bluetooth

2. Microsoft Excel

Dalam penggerjaan makalah ini, Microsoft excel digunakan untuk membuka dasli dengan format CSV untuk terlebih dahulu diidentifikasi serta membantu dalam penghitungan statistik lainnya.

3. Julia (*Programming Language*) versi 1.6.2 dengan menggunakan paket berikut :

Nama Paket	Fungsi
CSV.jl	Menangani data teks dengan format delimited, baik itu comma-delimited (csv), tab-delimited (tsv), atau lainnya agar dapat terbaca.
DataFrames.jl	Menyediakan seperangkat <i>tools</i> untuk bekerja dengan data berbentuk tabel di Julia. Desain dan fungsinya mirip dengan pandas (dalam Python) dan data.frame, data.table dan dplyr (dalam R).
Plots.jl	<i>Toolset</i> untuk keperluan visualisasi. Bekerja sama dengan backend lain, seperti GR, PyPlot, PGFPlotsX, atau Plotly.
Measures.jl	Memungkinkan untuk membangun dan menampilkan berbagai jenis unit atau satuan dan kemudian dapat dievaluasi.
VMLS.jl	Mengaktifkan fungsi berupa vektor, matriks dan <i>least square</i> .
LinearAlgebra.jl	Memungkinkan pengoperasian berbasis aritmatika array, faktorisasi matriks dan fungsi terkait aljabar linier dapat dijalankan.

2.2. Data

Data yang digunakan adalah Data Curah Hujan 15 Kota di Kalimantan, yaitu Samarinda, Balikpapan, Sendawar, Tarakan, Tanjung Selor, Malinau, Palangkaraya, Puruk Cahu, Pangkalbanu, Banjarmasin, Kota Baru, Barabai, Pontianak, Ketapang, dan Sintang dari artikel <https://doi.org/10.1088/1742-6596/1796/1/012053>. Khusus untuk makalah ini, data yang diambil merupakan data Curah Hujan untuk Kota Samarinda, Tanjung Selor, Palangkaraya, Banjarmasin, dan Sintang. Data tersebut berjenis data *time series* karena curah hujan dipantau setiap bulan selama kurun waktu 19 tahun dengan masing-masing kota memuat 228 buah data .

2.3. Times Series Trend

Misalkan suatu data merepresentasikan suatu rangkaian dari sampel y pada suatu waktu(masa) $x^{(i)} = i$, maka dengan menggunakan *the straight line fit* untuk data runtun waktu dapat ditentukan *trend line* atau garis tren sebagai berikut :

$$\hat{y}^{(i)} = \theta_1 + \theta_2 i; \quad i = 1, 2, \dots, N, \quad (1)$$

Dengan θ_2 merupakan kemiringan pada setiap titik yang berada di sepanjang garis lurus $\hat{y}^{(i)}$ yang menginterpretasikan tren dari waktu ke waktu. Dengan mengurangi garis *trend* dari deret waktu asalnya, maka akan diperoleh *detrended time series*, $y^d - \hat{y}^d$. Selanjutnya, $y^d = (y^{(1)}, y^{(2)}, \dots, y^{(N)})$ dapat diaproksimasi dengan rumusan berikut :

$$y^d \approx \hat{y}^d = \hat{y}^{const} + \hat{y}^{lin}, \quad (2)$$

dengan

$$\hat{y}^{const} = \theta_1 \cdot 1, \quad \hat{y}^{lin} = \theta_2 \begin{pmatrix} 1 \\ 2 \\ \vdots \\ N \end{pmatrix}. \quad (3)$$

Substitusikan (3) ke (2), diperoleh sistem berikut:

$$\begin{aligned} y^{(1)} &\approx \hat{y}^{(1)} = \theta_1 \cdot 1 + \theta_2 \cdot 1 \\ y^{(2)} &\approx \hat{y}^{(2)} = \theta_1 \cdot 1 + \theta_2 \cdot 2 \\ &\vdots \\ y^{(N-1)} &\approx \hat{y}^{(N-1)} = \theta_1 \cdot 1 + \theta_2 \cdot (N-1) \\ y^{(N)} &\approx \hat{y}^{(N)} = \theta_1 \cdot 1 + \theta_2 \cdot N \end{aligned} \quad (4)$$

Sistem (4) dapat diselesaikan dengan menentukan solusi dari sistem persamaan linear berikut :

$$A\vec{\theta} = \hat{y}^d, \quad (5)$$

dengan

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & N \end{pmatrix}, \quad \vec{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \text{ dan } \hat{y}^d = \begin{pmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(N)} \end{pmatrix}.$$

Detrended time series sendiri menunjukkan kecocokan data tersebut dengan garis lurus $\hat{y}^{(i)}$. Ketika *detrended time series* bernilai positif maka time series berada di atas garis lurus $\hat{y}^{(i)}$, sebaliknya ketika bernilai negatif maka artinya time series berada di atas garis lurus $\hat{y}^{(i)}$ (Boyd *et al.* 2018).

2.4 Estimation Of Trend and Seasonal Component

Dalam aplikasinya, ada data *time series* dengan *trend* yang berulang secara berkala. Kondisi tersebut disebut dengan *trend* yang terpengaruh dengan pola musiman. Selanjutnya estimasi *trend* yang dipengaruhi oleh *seasonal* atau pola musiman tidak lagi dapat dikerjakan dengan menggunakan persamaan (5). Dengan demikian, estimasi *trend* dapat ditentukan sesuai dengan :

$$y^d \approx \hat{y}^d = \hat{y}^{lin} + \hat{y}^{seas}, \quad (6)$$

dengan

$$\hat{y}^{lin} = \theta_1 \begin{pmatrix} 1 \\ 2 \\ \vdots \\ N \end{pmatrix}, \quad \hat{y}^{seas} = \begin{pmatrix} \theta_2 \\ \theta_3 \\ \vdots \\ \theta_{(P+1)} \end{pmatrix}. \quad (7)$$

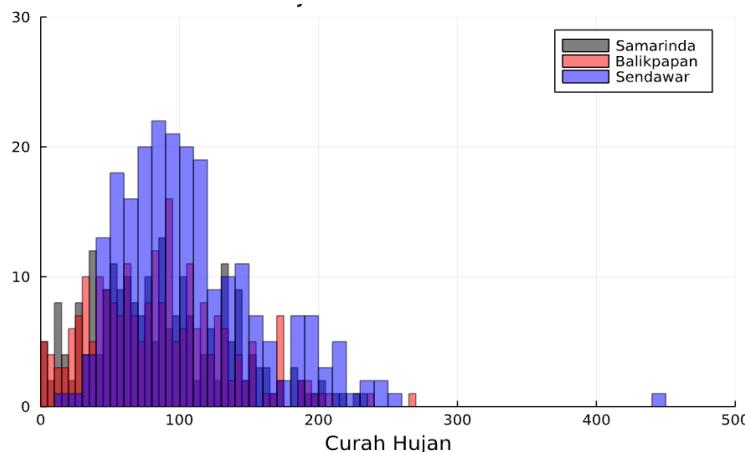
Komponen kedua bersifat periodik atau *seasonal*, dengan periode P (P dapat berupa P=12 yang menjelaskan jumlah bulan dalam satu tahun), dan terdiri dari $(\theta_2, \theta_3, \dots, \theta_{(P+1)})$ dengan perulangan sebanyak $\frac{N}{P}$ kali (dengan asumsi N adalah kelipatan dari P). Least Square fit dihitung dengan meminimalkan $\|A\theta - y^d\|^2$, dimana θ adalah vektor dengan ordo $(P+1) \times 1$. Selanjutnya, \hat{y}^d dapat diaproksimasi dengan SPL (Boyd *et al.* 2018),

$$\left(\begin{array}{cccc|c} 1 & 1 & 0 & \cdots & 0 \\ 2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P & 0 & 0 & \cdots & 1 \\ P+1 & 1 & 0 & \cdots & 0 \\ P+2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 2P & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ N-P+1 & 1 & 0 & \cdots & 0 \\ N-P+2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ N & 0 & 0 & \cdots & 1 \end{array} \right) \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \\ \theta_{(P+1)} \end{pmatrix} = \begin{pmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(N)} \end{pmatrix}. \quad (8)$$

3. Hasil dan Pembahasan

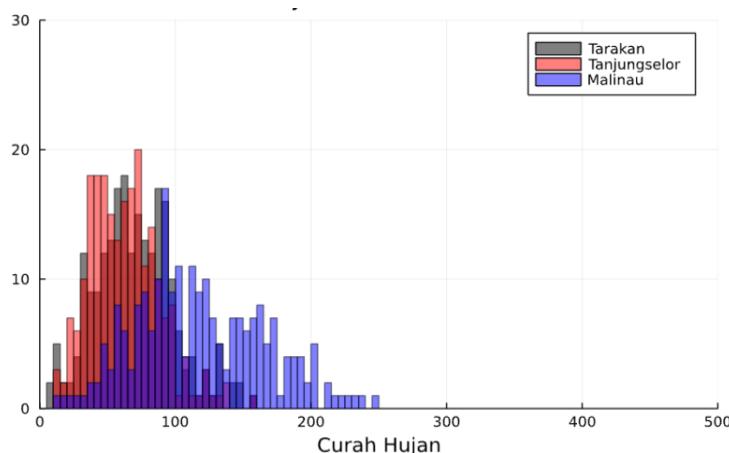
3.1 Eksplorasi Data

Dalam data curah hujan 15 kab/kota di Kalimantan yang terdiri dari Samarinda, Balikpapan, Sendawar, Tarakan, Tanjung Selor, Malinau, Palangkaraya, Puruk Cahu, Pangkalanbun, Banjarmasin, Kota Baru, Barabai, Pontianak, Ketapang, dan Sintang, penulis memperoleh bahwa :



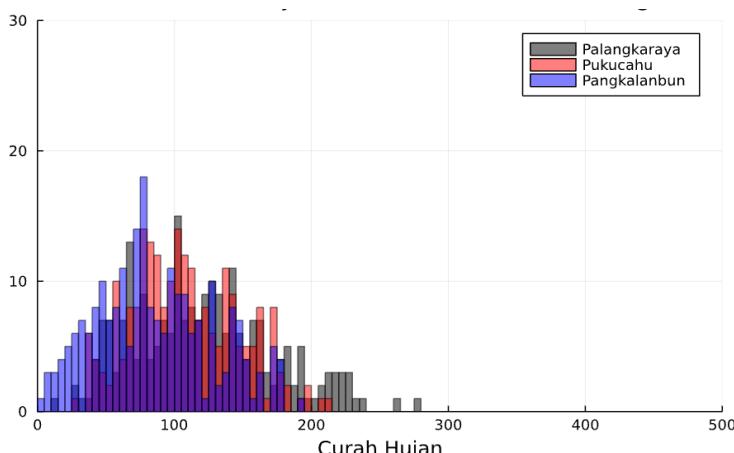
Gambar 1. Grafik Curah Hujan Provinsi Kalimantan Timur

Dalam histogram di atas dapat dilihat bahwa untuk Provinsi Kalimantan Timur yang diwakili Kota Samarinda, Balikpapan, dan Sendawar. Kota dengan curah hujan tertinggi pada Provinsi Kalimantan Timur ialah Kota Sendawar yang memiliki curah hujan disekitaran ≤ 200 mm dan mencapai frekuensi tertinggi disekitaran 100mm.



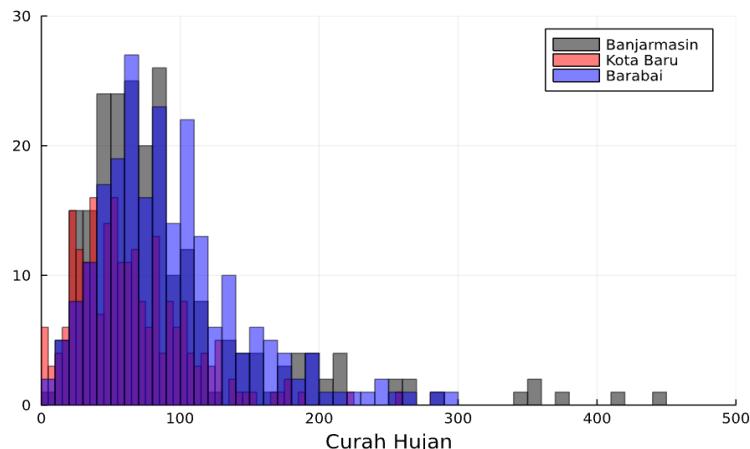
Gambar 2. Grafik Curah Hujan Provinsi Kalimantan Utara

Dalam histogram di atas dapat dilihat bahwa untuk Provinsi Kalimantan Timur yang diwakili Tarakan, Tanjung Selor, dan Malinau. Perhatikan bahwa ketiga kota cenderung memiliki curah hujan yang sama dan beriringan terutama pada selang 0-100 mm dengan frekuensi di sekitar ≤ 20 . Sementara untuk kota dengan curah hujan tertinggi dapat dilihat untuk curah hujan ≥ 200 mm hanya terdapat pada Kota Malinau.



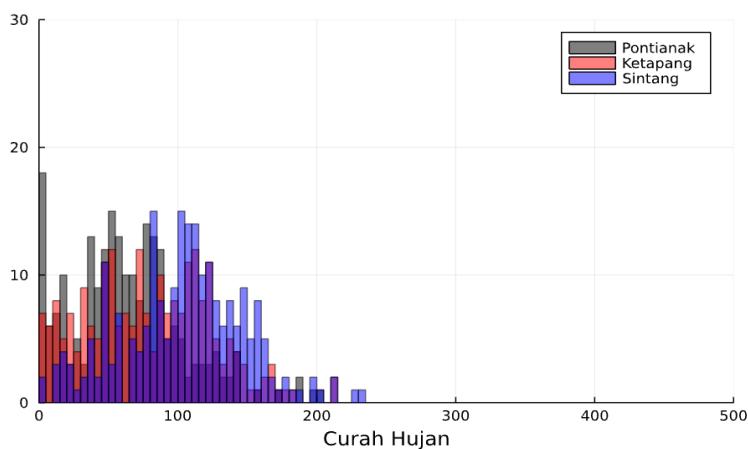
Gambar 3. Grafik Curah Hujan Provinsi Kalimantan Tengah

Dalam histogram di atas dapat dilihat bahwa untuk Provinsi Kalimantan Tengah yang diwakili Palangkaraya, Purukcahu, dan Pangkalanbun. Perhatikan bahwa ketiga kota cenderung memiliki curah hujan yang sama dan beriringan frekuensinya terutama pada selang 0-200 mm dengan frekuensi di sekitar ≤ 20 . Sementara untuk kota dengan curah hujan tertinggi dapat dilihat untuk curah hujan ≥ 200 mm hanya terdapat pada Kota Palangkaraya, namun tidak dengan frekuensi yang tinggi.



Gambar 4. Grafik Curah Hujan Provinsi Kalimantan Selatan

Dalam histogram di atas dapat dilihat bahwa untuk Provinsi Kalimantan Selatan yang diwakili Banjarmasin, Kota Baru, dan Barabai. Perhatikan bahwa ketiga kota cenderung memiliki curah hujan yang sama dan beriringan frekuensinya terutama pada selang 0-100 mm dengan frekuensi di sekitar ≤ 30 . Sementara untuk kota dengan curah hujan tertinggi dapat dilihat untuk curah hujan ≥ 300 mm hanya terdapat pada Kota Banjarmasin, namun tidak dengan frekuensi yang tinggi.

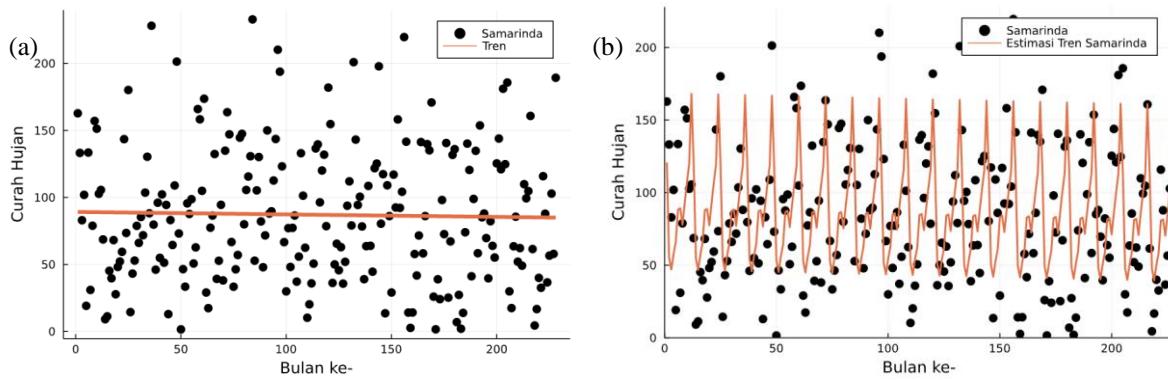


Gambar 5. Grafik Curah Hujan Provinsi Kalimantan Barat

Dalam histogram di atas dapat dilihat bahwa untuk Provinsi Kalimantan Barat yang diwakili Pontianak, Ketapang, dan Sintang. Perhatikan bahwa ketiga kota cenderung memiliki curah hujan yang sama dan beriringan frekuensinya terutama pada selang 0-200 mm dengan frekuensi di sekitar ≤ 20 . Sementara untuk kota dengan curah hujan tertinggi dapat dilihat untuk curah hujan ≥ 200 mm hanya terdapat pada Kota Sintang, namun tidak dengan frekuensi yang tinggi.

3.2 Estimasi Trend

Dalam makalah ini, dari data curah hujan di Kota Samarinda, Tanjung Selor, Palangkaraya, Banjarmasin, dan Sintang dalam kurun waktu 19 tahun di setiap bulannya, penulis memperoleh hasil analisis *trend* dari data tersebut apakah naik atau turun. Analisa tersebut berasal dari estimasi *trend* yang tidak terpengaruh pola musiman dan *trend* yang terpengaruh pola musiman. Analisa tersebut dapat terlihat dari visualisasi berikut ini.

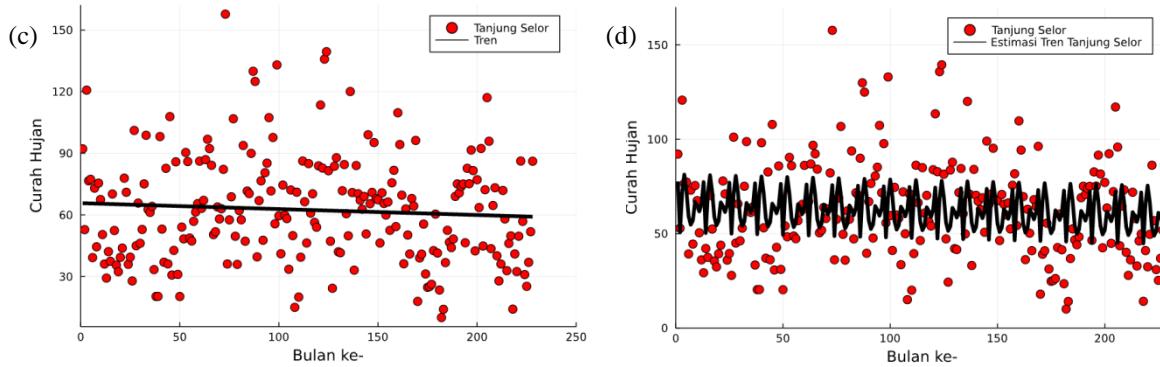


Gambar 6. Grafik estimasi trend curah hujan di kota Samarinda (a) tidak dipengaruhi pola musiman (b) dipengaruhi pola musiman .

Dengan memperhatikan grafik *trend* di atas, terlihat bahwa pada gambar (a) *trend* dari data curah hujan di Kota Samarinda turun. Hal ini, juga diperkuat dengan formula yang diperoleh, yaitu

$$\hat{y}^{(i)} = 89.18015461401026 - 0.01852848357252519i$$

Selanjutnya, dari gambar (b) terlihat bahwa *trend* dari data curah hujan di Kota Samarinda cenderung naik-turun atau terjadi gerakan musiman pada waktu yang sama karena pergantian musim dalam satu tahun.

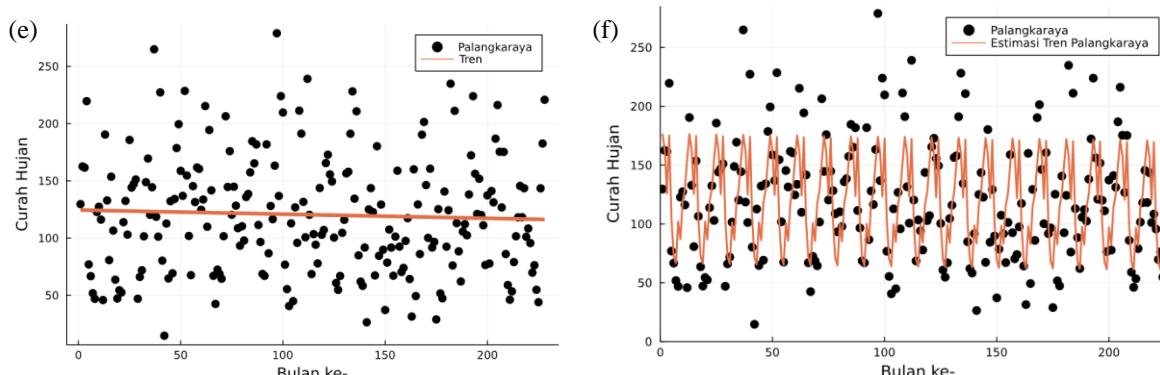


Gambar 7. Grafik estimasi trend curah hujan di kota Tanjung Selor (c) tidak dipengaruhi pola musiman (d) dipengaruhi pola musiman.

Dengan memperhatikan grafik *trend* di atas, terlihat bahwa pada gambar (c) *trend* dari data curah hujan di Kota Tanjung Selor turun. Hal ini, juga diperkuat dengan formula yang diperoleh, yaitu

$$\hat{y}^{(i)} = 65.7382784896043 - 0.02888246215620084i$$

Selanjutnya, dari gambar (d) terlihat bahwa *trend* dari data curah hujan di Kota Tanjung Selor cenderung naik-turun atau terjadi gerakan musiman pada waktu yang sama karena pergantian musim dalam satu tahun dan terjadi dua kali kenaikan *trend* dalam waktu yang berdekatan.

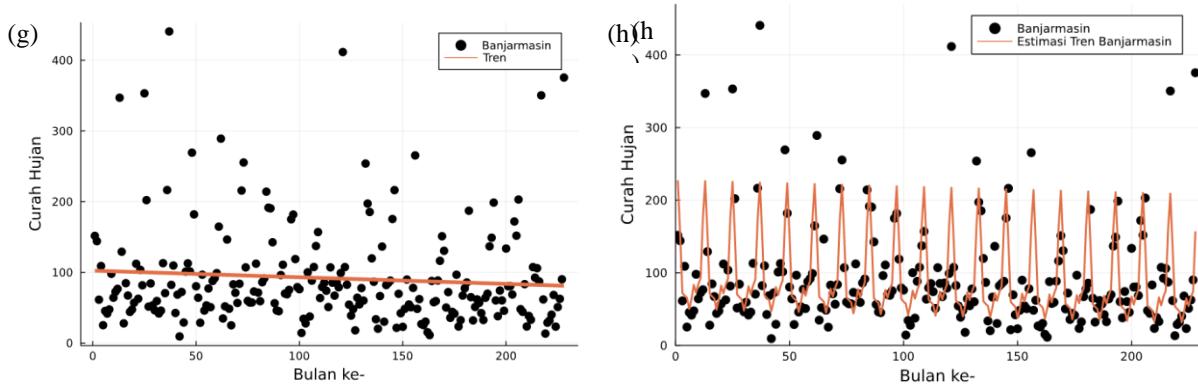


Gambar 8. Grafik estimasi trend curah hujan di kota Palangkaraya (e) tidak dipengaruhi pola musiman (f) dipengaruhi pola musiman

Dengan memperhatikan grafik *trend* di atas, terlihat bahwa pada gambar (e) *trend* dari data curah hujan di Kota Palangkaraya turun. Hal ini, juga diperkuat dengan formula yang diperoleh, yaitu

$$\hat{y}^{(i)} = 124.50819796173202 - 0.03611693173542087i$$

Selanjutnya, dari gambar (d) terlihat bahwa *trend* dari data curah hujan di Kota Palangkaraya cenderung naik-turun atau terjadi gerakan musiman pada waktu yang sama karena pergantian musim dalam satu tahun dan terjadi dua kali kenaikan *trend* dalam waktu yang berdekatan.

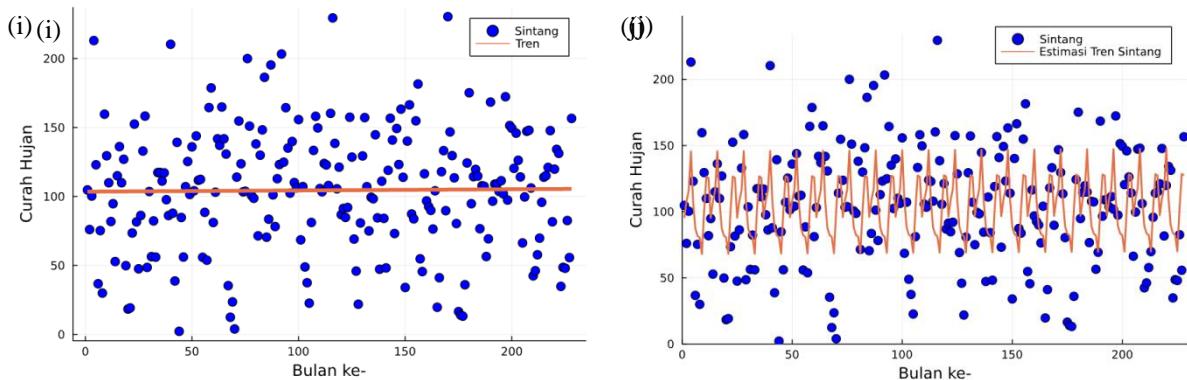


Gambar 9. Grafik estimasi *trend* curah hujan di kota Banjarmasin (g) tidak dipengaruhi pola musiman (h) dipengaruhi pola musiman.

Dengan memperhatikan grafik *trend* di atas, terlihat bahwa pada gambar (g) *trend* dari data curah hujan di Kota Banjarmasin turun. Hal ini, juga diperkuat dengan formula yang diperoleh, yaitu

$$\hat{y}^{(i)} = 1102.47450180041012 - 0.0938719806807061i$$

Selanjutnya, dari gambar (h) terlihat bahwa *trend* dari data curah hujan di Kota Banjarmasin cenderung naik-turun atau terjadi gerakan musiman pada waktu yang sama karena pergantian musim dalam satu tahun.



Gambar 10. Grafik estimasi *trend* curah hujan di kota Sintang (i) tidak dipengaruhi pola musiman (j) dipengaruhi pola musiman .

Dengan memperhatikan grafik *trend* di atas, terlihat bahwa pada gambar (i) *trend* dari data curah hujan di Kota Sintang naik. Hal ini, juga diperkuat dengan formula yang diperoleh, yaitu

$$\hat{y}^{(i)} = 103.5108627398675 + 0.008602803275002454i$$

Selanjutnya, dari gambar (j) terlihat bahwa *trend* dari data curah hujan di Kota Palangkaraya cenderung naik-turun atau terjadi gerakan musiman pada waktu yang sama karena pergantian musim dalam satu tahun dan terjadi dua kali kenaikan *trend* dalam waktu yang berdekatan.

Dari penelitian sebelumnya ditentukan pola curah hujan dengan ditunjang oleh FFT dan fungsi EOF dihasilkan bahwa curah hujan di Kalimantan terdiri dari 2 pola yaitu monsun(Balikpapan, Kotabaru, Palangkaraya, Barabai, Puruk Cahu, dan Banjarmasin) dan ekuatorial(meliputi: Samarinda, Sendawar, Tarakan, Tanjung Selor, Pontianak, Ketapang, Malinau, Pangkalanbun, dan Sintang) (Nurdiani *et al.* 2021). Sebagai perbandingan, dari makalah ini, dengan memperhatikan *trend* yang dipengaruhi pola musiman diperoleh 2 pola curah hujan yaitu monsun dimana *trend* curah hujan bergerak naik turun dengan pola tetap(Samarinda dan Banjarmasin), serta pola ekuatorial dimana *trend* curah hujan

cenderung naik-turun atau terjadi gerakan musiman pada waktu yang sama karena pergantian musim dalam satu tahun dan terjadi dua kali kenaikan *trend* dalam waktu yang berdekatan (Tanjung Selor, Palangkaraya, dan Sintang).

Untuk mengetahui metode estimasi mana yang lebih cocok digunakan untuk data curah hujan beberapa kota di Kalimantan ini dibutuhkan RMSE, berikut disajikan data nilai RMSE masing-masing kota :

Tabel 1. Tabel Standar Deviasi dan RMSE

Kota	Standar Deviasi	Roots Mean Square (RMSE)	
		Estimasi <i>trend</i> yang tidak dipengaruhi pola musiman	Estimasi <i>trend</i> yang dipengaruhi pola musiman
Samarinda	50.2465	50.2317	37.8239
Tanjung Selor	25.8299	25.7599	24.1824
Palangkaraya	51.1800	51.1248	35.7094
Banjarmasin	71.1639	70.8952	47.8565
Sintang	44.2272	44.2236	38.6163

Standar deviasi dapat digunakan untuk menggambarkan seberapa besar keragaman dari data, dimana semakin kecil nilainya maka keragaman data semakin homogen. Dari tabel 1, dapat dilihat bahwa data curah hujan di Tanjung Selor merupakan data dengan standar deviasi terkecil. Selanjutnya, dari RMSE dapat dilihat bahwa RMSE dari estimasi *trend* yang tidak dipengaruhi pola musiman cenderung lebih besar dibandingkan RMSE dari estimasi *trend* yang dipengaruhi pola musiman, sehingga estimasi *trend* yang dipengaruhi pola musiman lebih cocok diterapkan untuk mengestimasi *trend* dari data curah hujan dari beberapa kota di Kalimantan.

4. Simpulan

Telah dilakukan dua metode estimasi yaitu estimasi *trend* yang tidak dipengaruhi oleh pola musiman dan estimasi *trend* yang dipengaruhi oleh pola musiman. Dari estimasi *trend* data curah hujan yang tidak dipengaruhi oleh pola musiman diperoleh bahwa 4 Kota yaitu Samarinda, Tanjung Selor, Palangkaraya, dan Banjarmasin dari 5 kota yang penulis pilih memiliki *trend* turun, serta Sintang menjadi satu-satunya kota dengan *trend* data curah hujan naik. Oleh karena itu dapat disimpulkan untuk data curah hujan beberapa kota di Kalimantan yang dipilih penulis, dari persamaan (1), jika θ_2 bernilai negatif, maka *trend* data turun. Selanjutnya untuk estimasi *trend* data yang dipengaruhi oleh pola musiman diperoleh bahwa *trend* dari data curah hujan cenderung naik-turun atau terjadi gerakan musiman pada waktu yang sama karena pergantian musim dalam satu tahun (pola curah hujan munson) pada kota Samarinda dan Banjarmasin. Disisi lain terjadi gerakan musiman dan dua kali kenaikan *trend* dalam waktu yang berdekatan (pola curah hujan ekuatorial) pada Kota Tanjung Selor, Palangkaraya, dan Sintang.

Untuk perhitungan standar deviasi dan *Root Mean Square Error* (RMSE) diperoleh data curah hujan di Tanjung Selor merupakan data dengan standar deviasi terkecil. Selanjutnya, dari RMSE dapat dilihat bahwa RMSE dari estimasi *trend* yang tidak dipengaruhi pola musiman cenderung lebih besar dibandingkan RMSE dari estimasi *Trend* yang dipengaruhi pola musiman, sehingga estimasi *trend* yang dipengaruhi pola musiman lebih cocok diterapkan untuk mengestimasi *trend* dari data curah hujan dari beberapa kota di Kalimantan.

5. Ucapan Terima Kasih

Penulis ingin mengucapkan terima kasih kepada Moh. Khoirun Najib selaku asisten mata kuliah Komputasi Matematika atas ilmu yang telah diberikan selama proses praktikum dan atas saran serta kritik yang membangun. Seterusnya, dalam penulisan makalah ini penulis mendapat banyak ide, saran serta kritik dari beberapa pihak dengan harapan agar makalah ini lebih baik. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada Fitra Nuvus Salsabila, Trianty P. Blante, Anggun Praptaningsih, dan Ade Irawan atas kontribusi selama proses pembuatan makalah ini.

6. Daftar Pustaka

- Harijono S W B, Juaeni I, Lubis A, Ruminta dan Tyasyono, 2010, Dampak Variasi Temperatur Samudera Pasifik dan Hindia Ekuatorial Terhadap Curah Hujan di Indonesia, *Jurnal Sains Dirgantara* 5 1–3.
- Ramage, C. S., 1971. *Monsoon Meteorology*. Academic Press, 269.
- Webster, P. J. 1987. *The Elementary Monsoon. Dalam Fein, J.S., dan Stephen, P. L., (eds). Monsoon*, New York : John Wiley and son.
- Tjasyono, B. H. K. 2006. Karakteristik dan Sirkulasi Atmosfer, BMG, Jakarta.
- Sutikno, Rokhana D.B, Putri S, dan Istriana. 2010. Prakiraan Cuaca dengan Metode Autoregressive Integrated Moving Average, Neural Network dan Adaptive Splines Threshold Autoregression di Stasiun Juanda Surabaya, *Jurnal Sains Dirgantara*, Vol. 8, No. 1: 43-61.

- Prasetya R. 2011. Analisis Curah Hujan Akibat Siklon Tropis Nangka, Parma dan Nida di Sulawesi Utara Skripsi. Manado: Universitas Sam Ratulangi.
- Satya Wardhana H, Yulihastin E. 2016. Interaksi El-Nino, Monsun dan Topografi Lokal Terhadap Anomali Curah Hujan Di Pulau Jawa
- Nurdiati S, Khatizah E, Najib MK, Hidayah RR. 2021. Analysis of rainfall patterns in Kalimantan using Fast Fourier Transform (FFT) and Empirical Orthogonal Function (EOF). *IOP Conf Ser Earth Environ Sci.* 1796(1):012053. doi:10.1088/1742-6596/1796/1/012053.
- Chang C-P, Wang Z, McBride J dan Liu C-H. 2005. *Annual cycle of Southeast Asia-Maritime Continent rainfall and the asymmetric monsoon transition* *Journal of Climate* 18 287–301
- Cryer Jonathan D. 1986. *Time Series Analysis*. Boston: Duxbury Press.
- S Boyd, L Vandenberghe. 2018. *Introduction to Applied Linear Algebra Vectors, Matrices, and Least Squares*. United Kingdom: Cambridge University Press

Perbandingan Hasil *Clustering* Curah Hujan di Wilayah Kalimantan dengan Metode K-Means dan K-Medoids

Adelia Christi

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: adelia_christiadelia@apps.ipb.ac.id

Abstrak.

Efek *global warming* berdampak pada perubahan iklim dan cuaca di Indonesia. Salah satu wilayah di Indonesia yang ikut merasakan dampak perubahan cuaca dan iklim adalah Pulau Kalimantan. Sejumlah wilayah di Provinsi Kalimantan mengalami bencana banjir sepanjang tahun 2021. Oleh karena itu, curah hujan merupakan data yang penting untuk diketahui oleh masyarakat dan lembaga-lembaga terkait karena dapat memberikan informasi tentang pola curah hujan disuatu wilayah pada periode waktu tertentu. Intensitas curah hujan dapat dipengaruhi oleh beberapa faktor, seperti posisi lintang, ketinggian tempat, pola angin, sebaran bentang darat dan perairan, dll. Beberapa metode yang dapat digunakan adalah K-means dan K-medoids clustering. Berdasarkan proses clustering rata-rata curah hujan bulanan menjadi tiga cluster, yaitu curah hujan rendah, sedang, dan tinggi yang telah dilakukan dengan metode K-means dan K-medoids dapat disimpulkan bahwa terdapat perbedaan yang cukup mencolok di beberapa bulan tertentu, seperti bulan Januari, Maret, Mei, Juni, Juli, Oktober, dan Desember. Hal ini dapat terjadi karena adanya perbedaan cara iterasi. Kota-kota yang perlu diwaspadai setiap bulannya adalah kota yang tergolong sebagai kota dengan curah hujan yang tinggi, yaitu kota yang masuk ke cluster ke-3.

Kata Kunci: clustering, curah hujan, k-means, k-medoids.

1. Pendahuluan

Hujan adalah sebuah peristiwa presipitasi, yaitu jatuhnya butir-butir air ke permukaan bumi akibat terjadinya proses kondensasi uap air di atmosfer (Winarno *et al.* 2019). Curah hujan diukur sebagai tinggi air yang jatuh diperlakukan bumi yang datar dalam periode waktu tertentu. Menurut Badan Meteorologi Klimatologi dan Geofisika (BMKG), curah hujan (mm) adalah ketinggian air hujan yang terkumpul dalam penakar hujan pada tempat yang datar, tidak menyerap, tidak meresap dan tidak mengalir. Unsur hujan 1 (satu) milimeter artinya dalam luasan satu meter persegi pada tempat yang datar tertampung air hujan setinggi satu milimeter atau tertampung air hujan sebanyak satu liter. Berdasarkan besar curah hujan, BMKG membagi hujan menjadi tiga kategori, yaitu hujan sedang (20-50 mm per hari), hujan lebat (50-100 mm per hari), dan hujan sangat lebat (di atas 100 mm per hari). Intensitas curah hujan dapat dipengaruhi oleh beberapa faktor, seperti posisi lintang, ketinggian tempat, pola angin, sebaran bentang darat dan perairan, serta pegunungan dan gunung-gunung yang tinggi (Tukidin 2010). Secara geografis, Indonesia terletak di antara Benua Australia dan Asia, serta di antara Samudra Hindia dan Samudra Pasifik. Letak geografis ini mempunyai dampak angin laut yang membawa banyak hujan (Julisman 2013). Hal ini pula yang menyebabkan Indonesia memiliki iklim tropis. Selain itu, letak geografis Indonesia juga menyebabkan Indonesia memiliki dua musim. Hal ini dipengaruhi oleh angin musim yang berhembus tiap enam bulan sekali. Namun, pola musim tersebut kini tidak lagi dapat digunakan sebagai acuan karena efek *global warming* berdampak pada perubahan iklim dan cuaca di Indonesia (Fauziyah *et al.* 2013). Beberapa dampak yang ditimbulkan antara lain, musim di Indonesia berubah menjadi tidak menentu dan intensitas hujan meningkat. Intensitas curah hujan yang tinggi seringkali menjadi penyebab bencana banjir karena volume air yang berlebihan tidak dapat diserap oleh tanah.

Salah satu wilayah di Indonesia yang ikut merasakan dampak perubahan cuaca dan iklim adalah Pulau Kalimantan. Kalimantan merupakan pulau terbesar di Indonesia. Data Badan Nasional Penanggulangan Bencana (BNPB) sepanjang tahun 2021 memperlihatkan bahwa di wilayah Kalimantan terdapat tiga provinsi mengalami bencana banjir yang dipengaruhi oleh fenomena hidrometeorologi basah (curah hujan tinggi). Kejadian banjir di Kalimantan Tengah tercatat sebanyak tujuh kali, Kalimantan Barat sebanyak delapan kali, dan Kalimantan Selatan sebanyak empat kali. Variasi curah hujan yang tinggi belum ditunjang sarana observasi dan penanggulangan bencana yang memadai (Hasanah *et al.* 2017). Masih banyak lokasi-lokasi yang jarang atau bahkan tidak sama sekali memperoleh informasi mengenai cuaca dan iklimnya.

Berdasarkan uraian di atas, data curah hujan merupakan data yang penting untuk diketahui oleh masyarakat dan lembaga-lembaga terkait karena dapat memberikan informasi tentang pola curah hujan disuatu wilayah pada periode waktu tertentu. Oleh karena itu, pada tugas makalah ini, akan dilakukan pengelompokan kota-kota di wilayah Kalimantan berdasarkan data rata-rata curah hujan bulanan untuk memperoleh informasi kota-kota di Pulau Kalimantan yang perlu diwaspadai setiap bulannya apabila tergolong sebagai kota dengan curah hujan yang tinggi. Beberapa metode yang dapat digunakan adalah k-means dan k-medoids clustering. K-means clustering merupakan salah satu metode clustering yang

membagi n pengamatan menjadi k kelompok sedemikian sehingga tiap pengamatan termasuk ke dalam kelompok dengan rata-rata terdekat (Haraty *et al.* 2015). Sedangkan, K-medoids clustering adalah metode partisi karena menggunakan objek yang paling terpusat (*medoids*) di *cluster* menjadi pusat *cluster* dari nilai rata-rata objek dalam sebuah *cluster* (Marlina *et al.* 2018).

2. Data dan Metode

2.1. Alat dan Bahan

Dalam menyelesaikan tugas makalah ini, digunakan *personal computer* Lenovo Ideapad 330-14IGM. *Software* yang digunakan adalah Julia versi 1.6.2. Beberapa paket yang digunakan, di antara lain paket XLSX dan DataFrames untuk menyajikan data yang diperlukan untuk visualisasi dan analisis, paket Plots dan VegaLite digunakan untuk memvisualisasikan pola curah hujan di masing-masing kota, serta paket Clustering dan Distances yang digunakan untuk melakukan clustering dengan metode K-means dan K-medoids.

2.2. Data

Data yang digunakan dalam tugas ini adalah data curah hujan bulanan di 15 kota di lima provinsi di Kalimantan dari tahun 2001 sampai tahun 2019. Data curah hujan bulanan khususnya wilayah-wilayah di Indonesia dapat diperoleh dari situs dataonline.bmkg.go.id.

2.3. Metode

K-means merupakan metode pengklasteran secara *partitioning* yang memisahkan data ke dalam kelompok yang berbeda. Dengan *partitioning* secara iteratif, k-means mampu meminimalkan rata-rata jarak setiap data ke *cluster*-nya (Irwansyah dan Faisal 2015). Dalam algoritma k-means, setiap data harus termasuk ke *cluster* tertentu pada suatu tahapan proses, pada tahapan proses berikutnya dapat berpindah ke *cluster* yang lain. Dengan kata lain, data yang memiliki karakteristik sama dikelompokkan ke dalam satu *cluster* yang sama.

Metode K-means akan memilih pola k sebagai titik awal centroid secara acak. Jumlah iterasi untuk mencapai cluster centroid akan dipengaruhi oleh kandidat cluster centroid awal yang ditentukan secara acak dimana jika posisi centroid baru tidak berubah (Han *et al.* 2012). Nilai k yang dipilih menjadi centroid awal, akan dihitung dengan menggunakan rumus perhitungan jarak seperti Euclidean Distance, Manhattan, Cosine Similarity dan lain-lain. Metode perhitungan jarak adalah metode mencari jarak terdekat antara titik centroid dengan data. Data yang memiliki jarak terdekat dengan centroid akan membentuk sebuah cluster. Secara terperinci, algoritma k-means adalah sebagai berikut:

1. Tentukan k sebagai jumlah cluster yang akan dibentuk.
2. Tentukan k *centroid* awal secara acak
3. Hitung jarak setiap objek ke masing-masing centroid dari masing-masing cluster dengan menggunakan metode *Euclidian Distance* seperti pada persamaan (1).

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (1)$$

dengan d_{ik} adalah jarak antara data ke *centroid* dengan *centroid* ke- k , m adalah jumlah atribut, C_{ij} adalah data ke- i , c_k adalah data pusat klaster ke- k .

4. Alokasikan masing-masing objek ke dalam *centroid* yang paling dekat
5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan (2).

$$C = \frac{\Sigma^m}{n} \quad (2)$$

6. Ulangi langkah 3 jika posisi *centroid* baru tidak sama.

K-medoids merupakan suatu algoritma yang digunakan untuk menemukan *medoids* didalam sebuah *cluster* yang merupakan titik pusat dari suatu *cluster* (Sindi 2020). Langkah-langkah algoritma K-means adalah sebagai berikut:

1. Inisiasi pusat cluster sebanyak k .
2. Alokasikan setiap data ke *cluster* terdekat menggunakan persamaan ukuran jarak Euclidian Distance seperti pada persamaan (3).

$$d(x, y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (3)$$

3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat medoids yang baru.
4. Hitung jarak setiap objek yang berada masing-masing *cluster* dengan kandidat medoids baru.
5. Hitung total simpangan (S) dengan menghitung total *distance* baru – total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai medoid.

6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoids, sehingga diperoleh *cluster* beserta anggotanya masing-masing.

2.4. Tahapan

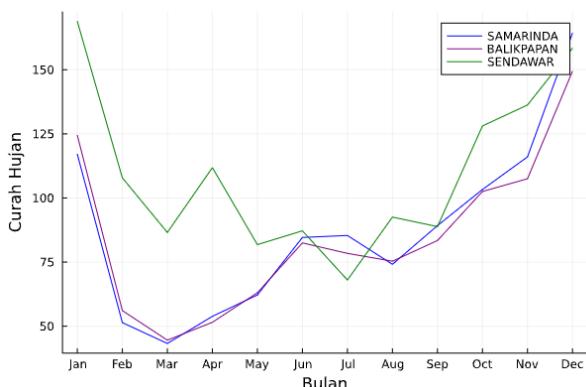
Tahapan-tahapan yang dilakukan dalam proses pengajaran makalah tugas ini secara garis besar terdiri dari tiga tahap, yaitu persiapan dan deskripsi data, visualisasi data, dan analisis data. Secara lengkap tahapan-tahapan yang dimaksud adalah sebagai berikut:

1. Mengolah data curah hujan di 15 kota di 5 provinsi di Kalimantan dari bulan Januari 2001 sampai bulan Desember 2019 dalam Microsoft Excel menjadi data rata-rata curah hujan bulanan.
2. Meng-import data dari Microsoft Excel ke Jupyter Notebook.
3. Memvisualisasikan data dengan menggunakan *line plot* untuk melihat pola rata-rata curah hujan bulanan di masing-masing kota.
4. Memvisualisasikan data rata-rata curah hujan bulanan di dua kota menggunakan *viplot* untuk melihat pola tebaran dua data yang memiliki atau tidak memiliki kemiripan pola rata-rata curah hujan bulanan.
5. Mengelompokkan kota-kota di Kalimantan berdasarkan data rata-rata curah hujan bulanan menggunakan metode *k-means* dan *k-medoids clustering* dengan membagi rata-rata curah hujan bulanan ke dalam tiga cluster, yaitu *cluster 1* untuk rata-rata curah hujan rendah, *cluster 2* untuk rata-rata curah hujan sedang, dan *cluster 3* untuk rata-rata curah hujan tinggi.
6. Menarik kesimpulan atau informasi kelompok kota-kota yang perlu diwaspadai oleh pemerintah daerah setempat setiap bulannya karena memiliki rata-rata curah hujan bulanan yang tinggi dibandingkan kelompok lain.

3. Hasil dan Pembahasan

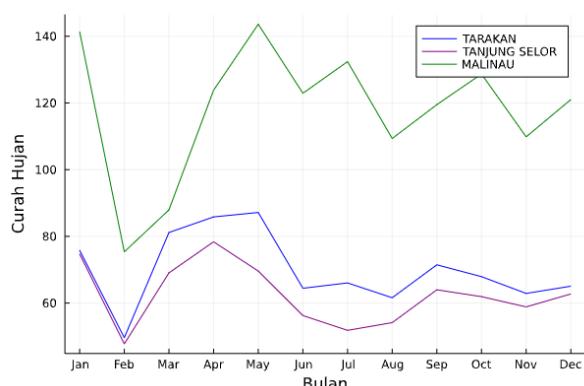
3.1. Visualisasi data

Hasil plot rata-rata curah hujan bulanan kota-kota di Provinsi Kalimantan Timur pada Gambar 1 menunjukkan bahwa ada kemiripan pola curah hujan di dua kota di Provinsi Kalimantan Timur, yaitu Samarinda dan Balikpapan. Curah hujan tertinggi terjadi pada bulan Desember dan terendah pada bulan Maret. Sedangkan, terdapat sedikit perbedaan pola curah hujan yang lebih fluktuatif di Kota Sendawar.



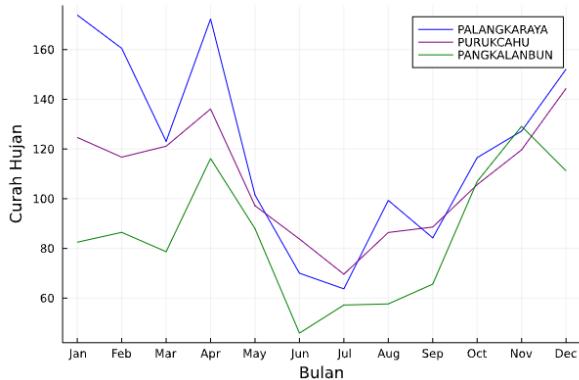
Gambar 1. Plot rata-rata curah hujan bulanan kota-kota di Provinsi Kalimantan Timur

Hasil plot rata-rata curah hujan bulanan kota-kota di Provinsi Kalimantan Utara pada Gambar 2 menunjukkan bahwa curah hujan tertinggi di provinsi ini terjadi pada sekitar bulan April dan Mei, sedangkan yang terendah pada bulan Februari. Kota yang memiliki pola curah hujan yang paling mirip adalah Kota Tarakan dan Tanjungselor.



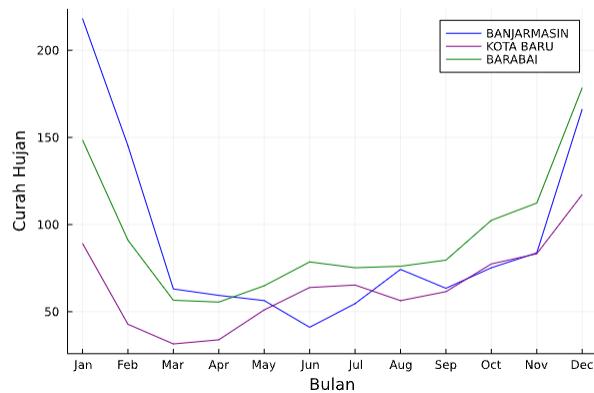
Gambar 2. Plot rata-rata curah hujan bulanan kota-kota di Provinsi Kalimantan Utara

Hasil plot rata-rata curah hujan bulanan kota-kota di Provinsi Kalimantan Tengah pada Gambar 3 menunjukkan bahwa curah hujan tertinggi di provinsi ini terjadi pada bulan April dan Desember, sedangkan yang terendah pada sekitar bulan Juni dan Juli.



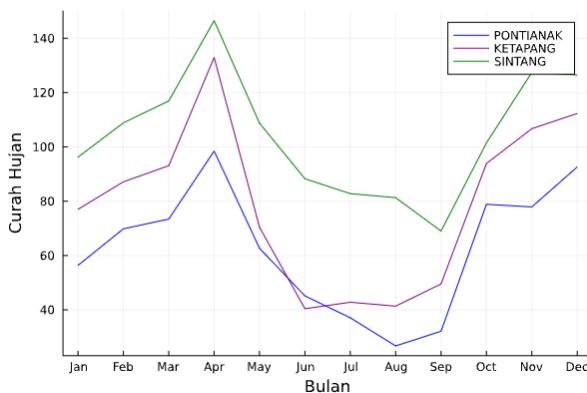
Gambar 3. Plot rata-rata curah hujan bulanan kota-kota di Provinsi Kalimantan Tengah

Hasil plot rata-rata curah hujan bulanan kota-kota di Provinsi Kalimantan Selatan pada Gambar 4 menunjukkan bahwa curah hujan tertinggi di provinsi ini terjadi pada sekitar bulan Desember dan Januari, sedangkan yang terendah pada sekitar bulan Maret dan April untuk Kota Baru dan Barabai, serta bulan Juni di Banjarmasin.



Gambar 4. Plot rata-rata curah hujan bulanan kota-kota di Provinsi Kalimantan Selatan

Hasil plot rata-rata curah hujan bulanan kota-kota di Provinsi Kalimantan Barat pada Gambar 5 menunjukkan bahwa curah hujan tertinggi di provinsi ini terjadi pada sekitar bulan April dan Desember, sedangkan yang terendah pada sekitar bulan Agustus dan September.



Gambar 5. Plot rata-rata curah hujan bulanan kota-kota di Provinsi Kalimantan Barat

3.2. Perbandingan kode untuk melakukan pengelompokan dengan metode k-means dan k-medoids clustering

Berikut adalah perbandingan kode pengelompokan kota-kota di wilayah Kalimantan berdasarkan data rata-rata curah hujan bulanan dengan metode k-means dan k-medoids clustering.

Kode 1. Contoh kode untuk memunculkan cluster dengan metode k-means clustering

```
using Clustering
#Cluster Bulan Januari dengan metode K-means
X1 = df_Cluster[:,2];
```

```

Y1 = convert.(Float64,X1[:,1])
C1 = kmeans(Y1', 3);
insertcols!(df_Cluster,3,:Cluster_Jan_Kmeans=>C1.assignments)

```

Kode 2. Contoh kode untuk memunculkan *cluster* dengan metode *k-medoids clustering*

```

using Distances
#Cluster Bulan Januari dengan metode K-medoids
D1 = pairwise(Euclidean(), Y1', Y1', dims=2);
K1 = kmedoids(D1, 3);
insertcols!(df_Cluster,4,:Cluster_Jan_Kmedoids=>K1.assignments)
select(df_Cluster, [:1, :2, :3, :4])

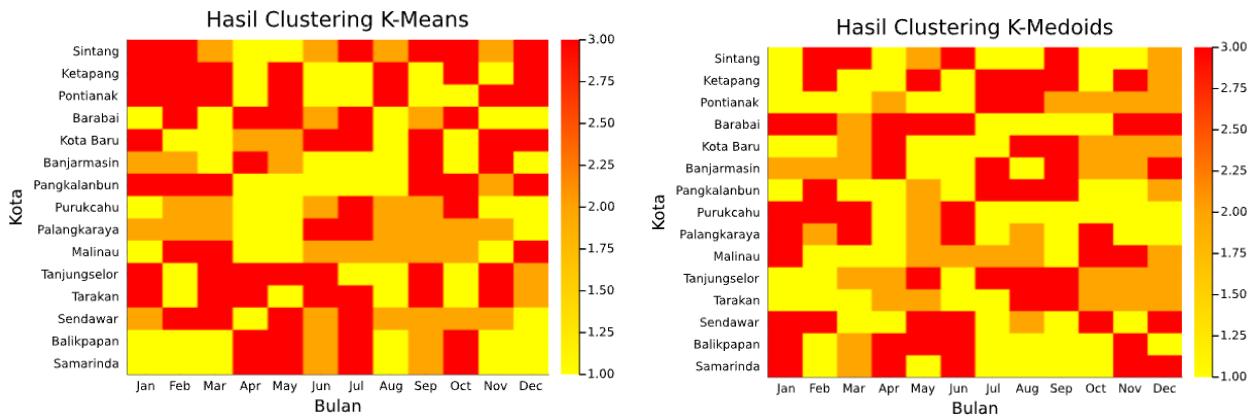
```

Pada Kode 1, X1 adalah kolom pada *dataframe* yang berisi data rata-rata curah hujan di 15 kota bulan Januari. Y1 adalah kolom data rata-rata curah hujan bulan Januari yang sudah di-*convert* tipe datanya menjadi *Float64*. Kemudian Y1 yang sudah di-*transpose* dibentuk menjadi tiga kelompok dengan menggunakan perintah *kmeans* yang disimpan pada variabel C1. Selanjutnya, untuk memunculkan hasil *clustering* menjadi salah satu kolom pada *dataframe*, digunakan perintah *insertcols*.

Pada Kode 2, D1 adalah matriks jarak Euclid. K1 adalah hasil clustering dengan metode *k-medoids*. Apabila ingin memunculkan hasil *clustering* menjadi salah satu kolom pada *dataframe*, digunakan perintah *insertcols*. Kemudian, untuk menampilkan *dataframe* yang hanya terdiri dari kolom nama-nama kota, data rata-rata curah hujan bulan Januari, dan hasil *clustering* menggunakan kedua metode dapat digunakan perintah *select* dan diikuti dengan indeks kolom yang ingin ditampilkan. Kode ini berlaku untuk memunculkan *cluster* menggunakan kedua metode pada bulan-bulan berikutnya dengan mengganti indeks-indeks kolom yang sesuai.

3.3. Perbandingan hasil clustering dengan metode K-Means dan K-Medoids

Hasil *clustering* antara kedua metode terdapat perbedaan yang cukup mencolok di beberapa bulan tertentu, seperti bulan Januari, Maret, Mei, Juni, Juli, Oktober, dan Desember. Kota-kota yang perlu diwaspadai setiap bulannya adalah kota yang tergolong sebagai kota dengan curah hujan yang tinggi, yaitu kota yang masuk ke *cluster* ke-3 baik menggunakan metode *k-means* maupun *k-medoids*. Daftar-daftar kota tersebut dapat dilihat secara lebih ringkas pada Gambar 6 berikut.



Gambar 6. Hasil clustering dengan metode *k-means* dan *k-medoids*

4. Simpulan

Berdasarkan proses *clustering* rata-rata curah hujan bulanan menjadi tiga cluster, yaitu curah hujan rendah, sedang, dan tinggi yang telah dilakukan dengan metode *K-means* dan *K-medoids* dapat disimpulkan bahwa terdapat perbedaan yang cukup mencolok di beberapa bulan tertentu, yaitu pada bulan Januari, Maret, Mei, Juni, Juli, Oktober, dan Desember. Hal ini dapat terjadi karena adanya perbedaan cara iterasi. Kota-kota yang perlu diwaspadai setiap bulannya adalah kota yang tergolong sebagai kota dengan curah hujan yang tinggi, yaitu kota yang masuk ke *cluster* ke-3.

5. Ucapan Terima Kasih

Terima kasih kepada Euis Aprianti dan Dillah Rismawati yang telah membantu memberikan saran dan membantu memeriksa apabila terdapat kekeliruan dalam proses pembuatan *coding* pada bagian visualisasi maupun analisis data sehingga tugas makalah ini dapat terselesaikan dengan baik.

6. Daftar Pustaka

- BMKG. 2017. Probabilitik Curah Hujan 20 mm (tiap 24 jam). [diakses 2021 Okt 11]. <https://www.bmkg.go.id/cuaca/probabilistik-curah-hujan.bmkg>.
- BNPB. 2021. Review Bencana Juli 2021: Bertemunya Bencana Hidrometeorologi Basah dan Kering. [diakses 2021 Okt 11]. <https://www.bnppb.go.id/berita/review-bencana-juli-2021-bertemunya-bencana-hidrometeorologi-basah-dan-kering>.
- Fauziyah S, Sobriyah, Susilowati. 2013. Analisis Karakteristik dan Intensitas Hujan Kota Surakarta. *Jurnal Matriks Teknik Sipil*. 1(2): 82-89. doi: [10.20961/mateksi.v1i2.37551](https://doi.org/10.20961/mateksi.v1i2.37551).
- Han J, Kamber M, Pei J. 2012. *Data Mining: Concepts and Techniques*. Ed ke-3. Massachusetts (MA): Morgan Kaufmann.
- Haraty RA, Dimishkleh M, Masud M. 2015. An Enhanced K-Means Clustering Algorithm for Pattern Discovery in Healthcare Data. *International Journal of Distributed Sensor Networks*. (6): 1-11. doi: [10.1155/2015/615740](https://doi.org/10.1155/2015/615740).
- Hasanah N, Ugiarto M, Puspitasari N. 2017. Sistem Pengelompokan Curah Hujan Menggunakan Metode K-Means di Wilayah Kalimantan Timur. *Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi*. 2(2): 122-126.
- Irwansyah E, Faisal M. 2015. *Advanced Clustering*: Teori dan Aplikasi. Yogyakarta: Deepublish.
- Julismin. 2013. Dampak dan Perubahan Iklim di Indonesia. *Jurnal Geografi*. 5(1): 40-46.
- Marlina D, Putri NF, Fernando A, Ramadhan A. 2018. Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT*. 4(2): 64-71. doi: [10.24014/coreit.v4i2.4498](https://doi.org/10.24014/coreit.v4i2.4498).
- Sindi S, Ningse WRO, Sihombing IA, Ilmi F, Hartama D. 2020. Analisis Algoritma K-Medoids Clustering dalam Pengelompokan Penyebaran Covid-19 di Indonesia. *Jurnal Teknologi Informasi*. 4(1): 166-173.
- Tukidi. 2010. Karakter Curah Hujan di Indonesia. *Jurnal Geografi*. 7(2): 136-145.
- Winarno GJ, Harianto SP, Santoso T. 2019. Klimatologi Pertanian. Bandarlampung: Pusaka Media.

Analisis Curah Hujan Rencana Maksimum Kota Sendawar, Banjarmasin, dan Kotabaru Menggunakan Metode Gumbel

Evi Ardiyani

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: ardiyanievi@apps.ipb.ac.id

Abstrak.

Kehidupan manusia sangat dipengaruhi oleh unsur geografis. Iklim adalah salah satu unsur geografis yang paling penting. Salah satu parameter iklim yang berpengaruh adalah curah hujan. Curah hujan memiliki dua sifat, yakni sifat menguntungkan dan sifat merugikan. Tingginya curah hujan dapat menyebabkan bencana seperti bencana banjir. Bencana banjir dapat menyebabkan kerugian yang besar. Curah hujan rencana atau rancangan adalah curah hujan maksimum yang mungkin terjadi pada periode waktu tertentu misalnya pada 2,5,10 hingga 100 tahun mendatang. Perencanaan curah hujan maksimum pada periode T-tahun hingga 100 tahun sangat penting dala, seperti ketika ingin merencanakan proyek pada DAS (Daerah Aliran Sungai) curah hujan maksimum adalah aspek yang sangat penting agar proyek dapat diperhitungkan dengan matang sehingga bangunan yang akan dibangun dapat bertahan lama. Perencanaan curah hujan maksimum sangat bermanfaat untuk antisipasi banjir karena dapat diprediksi curah hujan tertinggi dapat terjadi. Perhitungan curah hujan rencana pada penelitian ini menggunakan metode gumbel dengan hasil curah hujan maksimum pada periode 100 tahun di kota Sendawar adalah 358.98, kota Banjarmasin adalah 400.33, dan kota Kotabaru adalah 248.92.

Kata Kunci: Curah Hujan, Curah Hujan Rencana, Metode Gumbel.

1. Pendahuluan

Indonesia merupakan negara tropis yang dilewati oleh garis khatulistiwa, menyebabkan Indonesia memiliki 2 musim yaitu musim kemarau dan penghujan yang sangat mempengaruhi kehidupan makhluk hidup di wilayah Indonesia. Indonesia disebut sebagai *Maritime Continent* karena wilayah Indonesia terdiri dari gugusan ribuan pulau besar maupun kecil yang dipisahkan oleh lautan dan Samudera (Isnoro *et al.* 2018). Penting untuk mengenali faktor kedua musim tersebut agar dapat beraktivitas dengan lancar. Musim kemarau dapat menyebabkan bencana kekeringan, sedangkan musim penghujan dapat mengakibatkan bencana banjir (Basuki *et al.* 2009). Perubahan intensitas, jumlah dan durasi curah hujan memiliki implikasi serius dengan dampak nyata di sektor-sektor seperti pertanian, pengelolaan sumber daya air dan pengendalian banjir (Alam *et al.* 2018).

Tidak adanya hujan yang berkepanjangan mengakibatkan musim kemarau yang berkepanjangan menyebabkan sektor-sektor tidak dapat berproduksi secara maksimal, sedangkan peningkatan curah hujan cenderung mendukung sektor-sektor berproduksi tetapi curah hujan yang berlebih dapat mengakibatkan banjir (Osei *et al.* 2021). Curah hujan adalah jumlah air hujan yang jatuh selama periode tertentu yang pengukurnya menggunakan satuan tinggi diatas permukaan tanah horizontal (Basuki *et al.* 2009). Bencana banjir dapat menyebabkan kerugian yang besar, oleh karena itu penting untuk mengantisipasi bencana banjir. Semakin tinggi curah hujan, maka peluang terjadi bencana banjir juga akan semakin besar (Upomo *et al.* 2016). Setiap tahunnya, dampak dari bencana banjir kian memburuk. Banjir tahunan dan dampak dari banjir tahunan dapat dikurangi dengan memprediksi curah hujan maksimum (Vivekanandan. 2017). Kejadian hujan merupakan proses stokastik, sehingga perlu sebuah Analisis dengan menggunakan teori probabilitas dan analisis frekuensi. Analisis frekuensi dapat memberikan solusi yang cocok untuk memprediksi hidroparameter meteorologi baik dalam ruang dan waktu, yang melibatkan estimasi probabilitas terjadinya suatu peristiwa tertentu (Bhagat. 2017). Kebutuhan akan adanya informasi prediksi curah hujan sangat diperlukan dalam berbagai sektor (Dian *et al.* 2013).

Penyebab utama dari banjir adalah tingginya curah hujan, dengan mengetahui curah hujan maksimum yang mungkin terjadi, dapat diantisipasi bencana banjir tersebut. Curah hujan rencana atau rancangan adalah curah hujan maksimum yang mungkin terjadi pada periode tertentu seperti 2,5,10, hingga 100 tahun Curah hujan rencana maksimum dan kedalaman banjir untuk berbagai periode dapat diperkirakan dengan analisis frekuensi seperti, distribusi normal, distribusi log-normal, distribusi log-pearsor III dan distribusi gumbel (Mukherjee. 2013). Masing-masing distribusi memiliki kelemahan dan kelebihan masing-masing. Menurut Basuki *et al* (2009), penentuan distribusi yang digunakan dapat ditentukan dari data yang tersedia. Distribusi Gumbel telah digunakan untuk memperkirakan curah hujan maksimum dalam 1 hari dan 1 jam di stasiun pengukur hujan Dhaulakuan (Vivekanandan, 2017) serta metode Gumbel sudah digunakan untuk memprediksi curah hujan maksimum dalam periode ulang dalam periode 1 sampai 5 hari (Sasireka *et al.* 2019). Metode gumbel dapat dipastikan sebagai distribusi yang lebih baik untuk analisis curah hujan dan berguna untuk memprediksi wilayah banjir (Okeke dan Ehirobo. 2017).

Dari uraian diatas, penelitian ini bertujuan untuk memperoleh curah hujan rencana atau rancangan maksimum yang mungkin terjadi pada suatu periode tertentu dalam kurun waktu 2, 5, 10, hingga 100 tahun dengan menggunakan metode gumbel. Hasil penelitian dapat bermanfaat untuk memprediksi curah hujan rencana maksimum sehingga dapat mengantisipasi bencana banjir di masa mendatang.

2. Data dan Metode

2.1. Alat dan Bahan

Penelitian ini menggunakan *personal computer* dengan spesifikasi IdeaPad Gaming 3 15IMH05. Processor Intel® Core™ i7-10750H CPU @ 2.60GHz 2.59 GHz dan *Installed RAM* sebesar 16 GB

Penelitian ini menggunakan *software* Excell dan Julia untuk membantu proses pengolahan data dan *plot* sebuah grafik. Paket yang digunakan pada *software* Julia untuk membantu proses perhitungan dan *plot* grafik diantaranya :

```
using Plots, StatsPlots, Statistics, BenchmarkTools, DataFrames, DelimitedFiles
using CSV, using Measures, XLSX, VegaLite, VegaDatasets, Dates, Measures
```

2.2. Data

Sumber data diperoleh dari data curah hujan bulanan 15 kota di 5 provinsi Kalimantan. Jenis data adalah data Sekunder, data diperoleh dari artikel *Analysis of rainfall patterns in Kalimantan using fast fourier transform (FFT) and empirical orthogonal function (EOF)* (Nurdiati. 2021). Data curah hujan dari 15 kota diantaranya kota Samarinda, Balikpapan, Sendawar, Tarakan, Tanjungselor, Malinau, Palangkaraya, Purukcahu, Banjarmasin, Kotabaru, Barabai, Pontianak, Ketapang, dan Sintang. Data curah hujan yang digunakan adalah data curah hujan bulanan selama 20 tahun dari 2001 hingga 2019.

2.3. Metode Aritmatika

Metode aritmatika atau rata-rata aljabar adalah metode paling sederhana untuk menghitung hujan rerata pada suatu daerah. Pengukuran yang dilakukan di beberapa stasiun dalam waktu yang bersamaan dijumlahkan kemudian dibagi dengan jumlah stasiun (Lashari *et al.* 2017) Hujan rerata dapat dihitung dengan rumus

$$P = \frac{p_1 + p_2 + p_3 + \cdots + p_n}{n} \quad (1)$$

dengan P adalah curah hujan rata-rata (mm/bulan), P_i curah hujan ke- i (mm/bulan) dan n banyak data.

2.4. Metode Gumbel

Metode Gumbel adalah metode yang digunakan untuk menghitung curah hujan rencana. Perumusan metode gumbel yaitu (Gulap *et al.* 2019):

$$X_T = P + s \times K \quad (2)$$

Keterangan :

X_T = Perkiraan nilai yang diharapkan terjadi dengan periode ulang T-tahunan

s = Standar Deviasi

K = Faktor probabilitas K, dengan persamaan:

$$K = \frac{Y_T - Y_n}{S_n} \quad (3)$$

Y_T = $-\ln(-\ln \frac{T-1}{T})$

T = Periode ulang

Y_n = *Reduced Mean* pada Tabel 1. (Suripin. 2004).

S_n = *Reduced Standard Deviation* pada Tabel 2. (Suripin. 2004).

Tabel 1. Reduce Mean (Y_n)

N	0	1	2	3	4	5	6	7	8	9
10	0.4952	0.4996	0.5035	0.5070	0.5100	0.5128	0.5157	0.5181	0.5202	0.5220
20	0.5236	0.5252	0.5268	0.5283	0.5296	0.5309	0.5320	0.5332	0.5343	0.5353
30	0.5362	0.5371	0.5380	0.5388	0.8396	0.5403	0.5410	0.5418	0.5424	0.5436
40	0.5436	0.5442	0.5448	0.5453	0.5458	0.5463	0.5468	0.5473	0.5477	0.5481
50	0.5436	0.5489	0.5493	0.5497	0.5501	0.5504	0.5508	0.5511	0.5515	0.5518
60	0.5521	0.5524	0.5527	0.5530	0.5533	0.5535	0.5538	0.5540	0.5543	0.5545
70	0.5548	0.5550	0.5552	0.5555	0.5557	0.5559	0.5561	0.5563	0.5565	0.5567
80	0.5569	0.5570	0.5572	0.5574	0.0558	0.5578	0.5580	0.5581	0.5583	0.5585
90	0.5586	0.5587	0.5589	0.5591	0.5592	0.5593	0.5595	0.5596	0.5598	0.5599
100	0.5600	0.5602	0.5603	0.5604	0.5606	0.5607	0.5608	0.5609	0.5610	0.5611

Tabel 2. Reduce Standard Deviation (S_n)

N	0	1	2	3	4	5	6	7	8	9
10	0.9496	0.9676	0.9833	0.971	1.0095	1.0206	1.0316	1.0411	1.0493	1.0565
20	1.0628	1.0696	1.0754	1.0811	1.0864	1.0915	1.0961	1.1004	1.1047	1.1080
30	1.1124	1.1159	1.1193	1.1226	1.1255	1.1285	1.1313	1.1339	1.1363	1.1388
40	1.1413	1.1436	1.1458	1.1480	1.1499	1.1519	1.1538	1.1557	1.1574	1.1590
50	1.1607	1.1623	1.1638	1.1658	1.1667	1.1681	1.1696	1.1708	1.1721	1.1734
60	1.1747	1.1759	1.1770	1.1782	1.1793	1.1803	1.1814	1.1824	1.1834	1.1844
70	1.1854	1.1863	1.1873	1.1881	1.1890	1.1898	1.1906	1.1915	1.1923	1.1930
80	1.1938	1.1945	1.1953	1.1959	1.1967	1.1973	1.1980	1.1987	1.1994	1.2001
90	1.2007	1.2013	1.2020	1.2026	1.2032	1.2038	1.2044	1.2049	1.2055	1.2060
100	1.2065	1.2069	1.2073	1.2077	1.2081	1.2084	1.2087	1.2090	1.2093	1.2096

2.5. Kemencengan/Skewness

Kemencengan (*skewness*) merupakan derajat ketidaksimetrisan atau dapat juga didefinisikan sebagai penyimpangan kesimetrisan dari suatu distribusi. Perhitungan tingkat kemencengan dapat menggunakan (Upomo *et al.* 2016):

$$Cs = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n (x_i - \bar{x})^3 \quad (4)$$

Keterangan :

n = Jumlah data

\bar{x} = Rata-rata data hujan (mm)

S = Standar Deviasi

X = data hujan (mm)

Cs = Kemencengan/*Skewness*

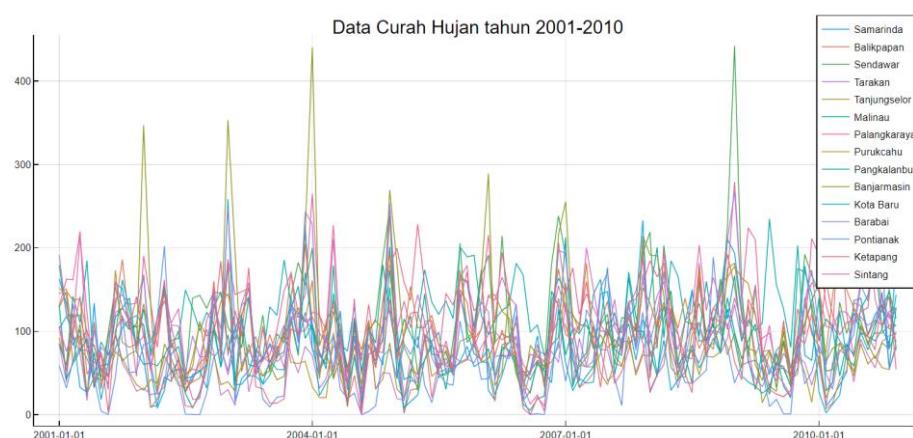
Tabel 3. Tabel syarat distribusi (Basuki *et al.* 2009)

Cs	Sebaran
>1.0	Gumbel
<1.0	Log Normal atau Log Pearson III
=1.0	Normal

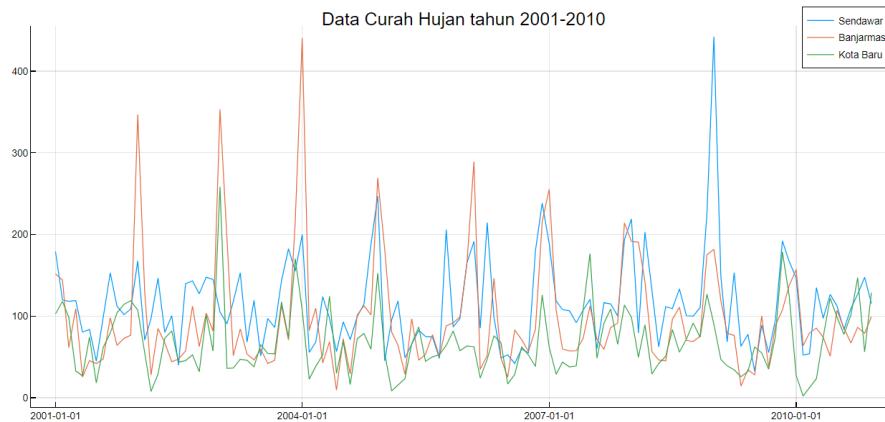
3. Hasil dan Pembahasan

3.1 Eksplorasi Data

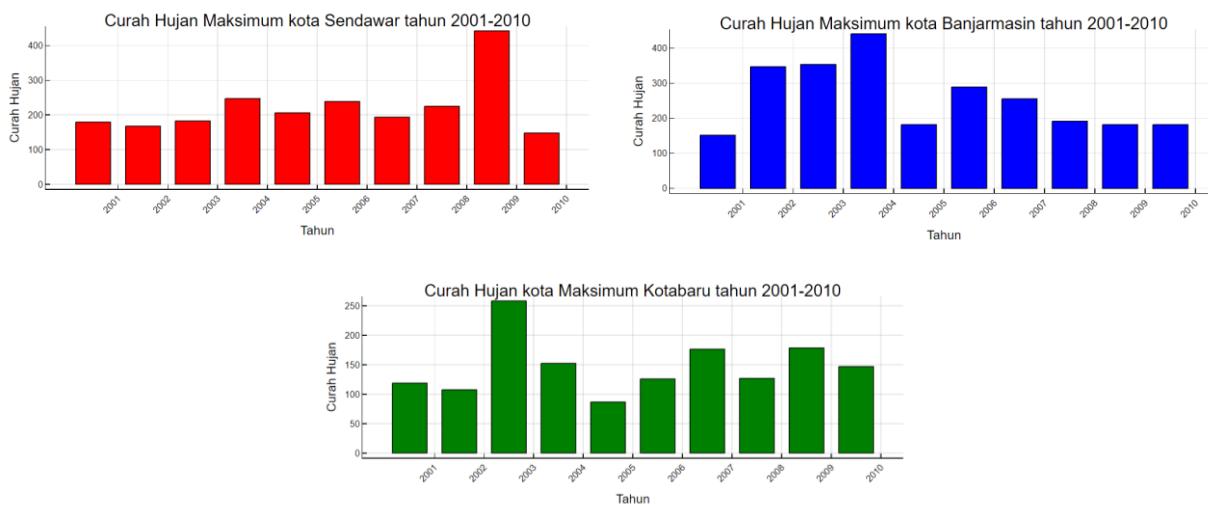
Data yang digunakan adalah data curah hujan 15 kota di provinsi Kalimantan (Nurdiati. 2021) yaitu kota, Samarinda, Balikpapan, Sendawar, Tarakan, Tanjungselor, Malinau, Palangkaraya, Purukcahu, Pangkalbanjar, Banjarmasin, Kotabaru, Barabai, Pontianak, Ketapang, dan Sintang. *Plot* data curah hujan pada 15 kota ditunjukkan pada Gambar. 1. Data yang diperoleh merupakan data curah hujan bulanan selama 10 tahun.

**Gambar 1.** Curah Hujan 15 Kota di Provinsi Kalimantan tahun 2001-2010

Data yang digunakan pada penelitian adalah data curah hujan 3 kota di provinsi Kalimantan, yaitu data curah hujan bulanan tahun 2001 sampai 2010 di kota Sendawar, Banjarmasin pada,dan Kotabaru pada Tabel 7. *Plot* data curah hujan bulanan ketiga kota pada Gambar 2. Pemilihan ketiga kota pada penilitian ini berdasarkan penentuan distribusi yang ditunjukkan oleh Gambar 4. Data ketiga kota pada tahun 2001-2010 digunakan untuk memprediksi curah hujan rencana maksimum pada periode-T tahun. Setiap tahun terdapat curah hujan maksimum yang terjadi pada Gambar 3 ditunjukkan curah hujan maksimum yang terjadi dalam periode satu tahun di ketiga kota.



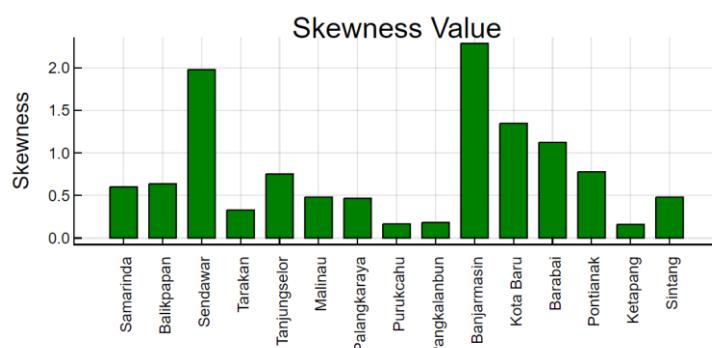
Gambar 2. Data Curah Hujan Kota Sendawar, Banjarmasin, dan Kotabaru tahun 2001 – 2010.



Gambar 3. Data Curah Hujan Maksimum Kota Sendawar, Banjarmasin, dan Kotabaru tahun 2001 – 2010.

3.2 Penentuan Distribusi

Perhitungan nilai C_s untuk data curah hujan ketiga kota dengan persamaan (1) menggunakan *software Julia* diperoleh hasil pada Gambar 4.



Gambar 4. Nilai Kemencenggan Data Curah Hujan 15 Kota Provinsi Kalimantan

Dari data diperoleh tiga nilai kemencengan tertinggi yaitu kota Sendawar, Banjarmasin, dan Kotabaru yang akan digunakan untuk menganalisis Curah Hujan Rencana periode T tahun.

Tabel 4. Tabel nilai C_s

Kota	C_s
Sendawar	1.58
Banjarmasin	2.26
Kotabaru	1.25

Kota Sendawar, Banjarmasin dan Kotabaru memiliki nilai $C_s > 1$, maka dapat dilanjutkan perhitungan curah hujan rencana maksimum dengan menggunakan metode Gumbel.

3.3 Metode Gumbel

Perhitungan curah hujan rencana menggunakan metode Gumbel dengan bantuan *software Julia*. Dalam perhitungan dengan menggunakan *software Julia*, dilakukan penurunan rumus untuk mempersingkat perhitungan agar diperoleh hasil dengan tingkat kesalahan yang sedikit. Dari persamaan (2), diperoleh

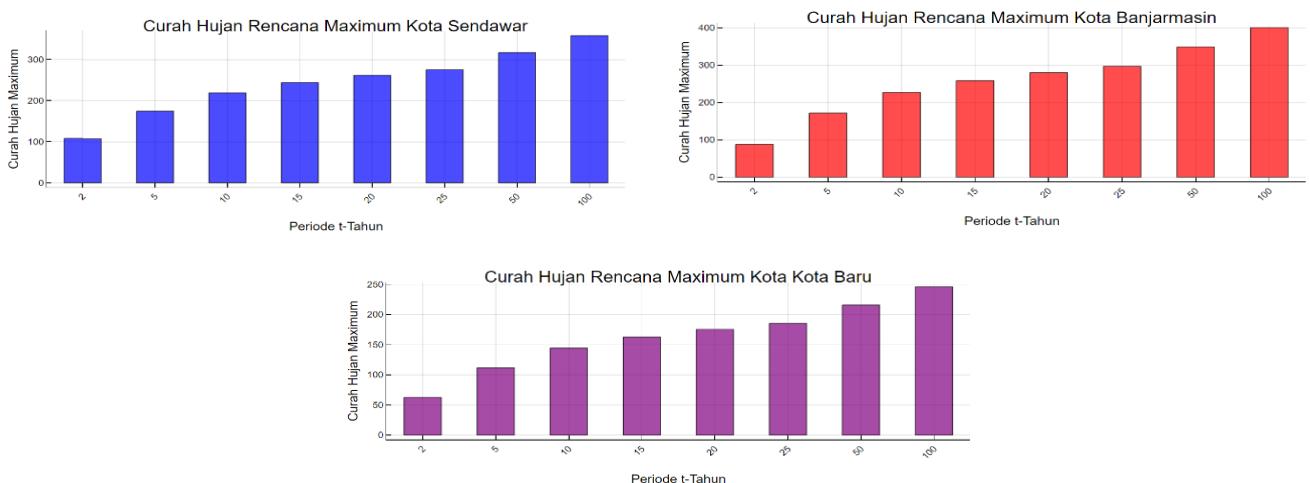
$$\begin{aligned} X_T &= P + s \times K = P + s \times \frac{Y_T - Y_n}{S_n} = P + s \times \frac{Y_T - Y_n}{S_n} \\ &= P + \frac{s}{S_n} Y_T - \frac{Y_n}{S_n} S = P - \frac{Y_n}{S_n} S + \frac{s}{S_n} Y_T \\ &= b + a \times Y_T \end{aligned} \quad (5)$$

dengan $b = P - \frac{Y_n}{S_n} S$, $a = \frac{s}{S_n}$

Hasil perhitungan curah hujan rencana menggunakan metode Gumbel ditunjukkan oleh Tabel 6, dengan $n = 10$. Berdasarkan Tabel 1, maka nilai Y_n yang digunakan sebesar 0.4952 dan nilai S_n sebesar 0.9496. Nilai Y_T pada Tabel 6 diperoleh dari persamaan (5). Hasil perhitungan curah hujan rencana menggunakan metode Gumbel pada kota Sendawar, Banjarmasin, dan Kotabaru ditunjukkan pada Gambar. 5.

Tabel 6. Tabel nilai X_T dan Y_T

Periode (T-tahun)	X_T Sendawar	X_T Banjarmasin	X_T Kotabaru	Y_T
2	108.06	81.65	62.59	0.366
5	175.24	171.36	111.67	1.499
10	219.71	226.79	144.17	2.250
15	224.81	258.06	162.50	2.673
20	262.38	279.95	175.34	2.970
25	275.91	296.81	185.23	3.198
50	317.60	348.76	215.69	3.901
100	358.98	400.33	248.92	4.600



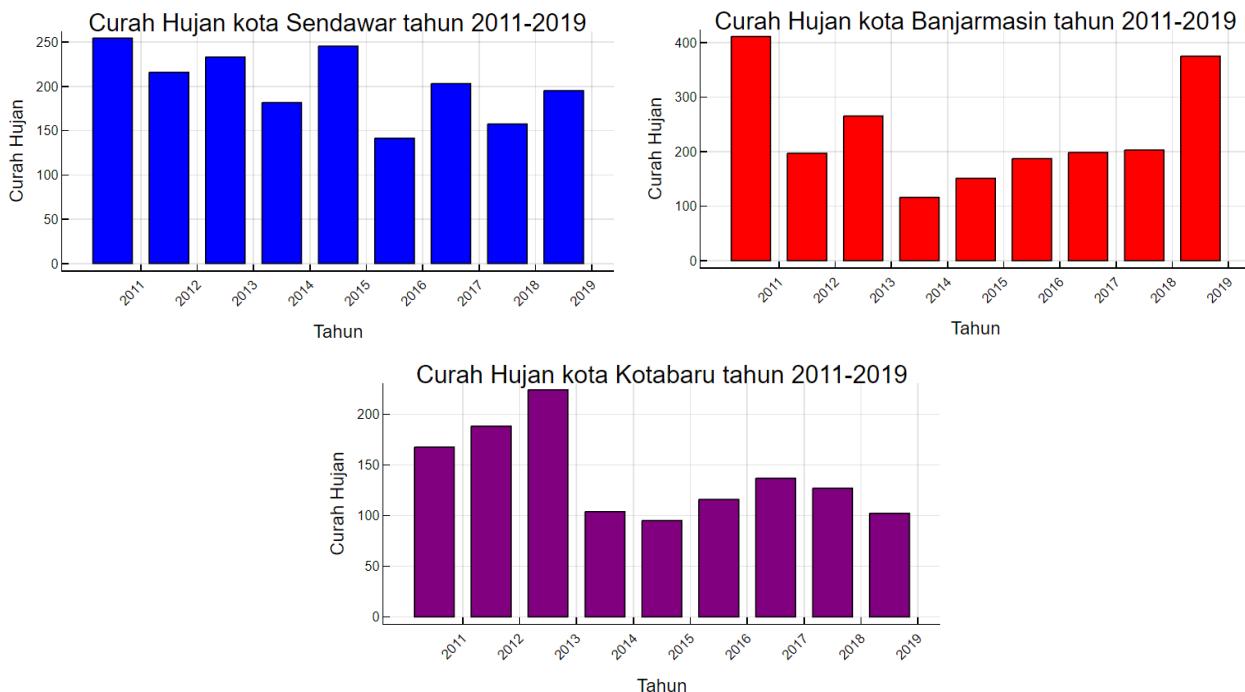
Gambar 5. Curah Hujan Rencana Kota Sendawar, Banjarmasin, dan Kotabaru

4. Simpulan

Dari hasil perhitungan curah hujan rencana maksimum dengan menggunakan metode Gumbel pada 3 kota di provinsi Kalimantan, yaitu Sendawar, Banjarmasin, dan Kotabaru dapat disimpulkan sebagai berikut :

1. Curah hujan maksimum kota Sendawar pada periode 2 tahun adalah 108.06 hingga periode 100 tahun adalah 358.98
2. Curah hujan maksimum kota Banjarmasin pada periode 2 tahun adalah 81.65 hingga periode 100 tahun adalah 400.33
3. Curah hujan maksimum kota Kotabaru pada periode 2 tahun adalah 62.59 hingga periode 100 tahun adalah 248.92

Peneliti membandingkan hasil perhitungan curah hujan rencana maksimum periode 2 tahun hingga 10 tahun menggunakan data curah hujan bulanan pada tahun 2001 – 2010 dengan data curah hujan maksimum pada tahun 2011 – 2019 yang ditunjukkan pada Gambar 6. Terdapat perbedaan pada prediksi curah hujan rencana maksimum kota Sendawar, Banjarmasin, dan Kotabaru dengan data curah hujan yang terjadi pada ketiga kota dalam kurun waktu 2011 – 2019. Perbedaan prediksi dengan kondisi yang sebenarnya terjadi dapat dipengaruhi oleh beberapa faktor, diantaranya jumlah data yang digunakan, pemilihan metode, dan nilai kesalahan yang terjadi (Satyarini. 2007).



Gambar 6. Curah Hujan Maksimum Kota Sendawar, Banjarmasin, dan Kotabaru tahun 2011 – 2019.

5. Daftar Pustaka

- Alam A M, Emura K, Farnham C, Yuan J. 2018. *Best-Fit Probability Distributions and Return Periods for Maximum Monthly Rainfall in Bangladesh. Department of housing and Environment Design, Climate* 6 : 9
- Basuki, Winarsih I, Adhyani N L. 2009. Analisis Periode Ulang Hujan Maksimum dengan Berbagai Metode. Badan Meteorologi, Klimatologi, dan Geofisika
- Bhagat N. 2017. *Flood Frequency Analysis Using Gumbel's Distribution Method : A Case Study of Lower Mahi Basin, India. Journal of Water Resources and Ocean Science*. Vol 6(4) : 51-54
- Isnoro K F N, Firdianto P U, Susilawati A. 2018. Studi Tentang Fenomena Borneo Vortex Terhadap Variabilitas Awan di Kalimantan Barat (Studi Kasus : 11-13 Januari 2018). *Jurnal Ilmu dan Inovasi Fisika*. Vol 2(02) : 127 - 136
- Lashari, Kusumardani I, Prakasa F. 2017. Analisis Distribusi Curah Hujan di Area Merapi Menggunakan Metode Aritmatika dan Poligon. *UNNES Journals*.
- Mukherjee M K. 2013. *Flood Frequency Analysis of River Subernarekha, India, Using Gumbel's Extreme Value Distribution. International Journal of Computational Engineering Research*. Vol 3 : 17
- Nurdiati S. 2021. *Analysis of rainfall patterns in Kalimantan using fast fourier transform (FFT) and empirical orthogonal function (EOF)*. IPB University.
- Okeke O B dan Ehiorobo JO. 2017. *Frequency Analysis of Rainfall for Flood Control in Patani, Delta State of Nigeria. Nigerian Journal of Technology (NIJOTECH)*. Vol 36(1) : 282 – 289
- Osei M A, Amekudzi L K, Omari-Sasu A Y, Yamba E I, Quansah E, Aryee J N A, Kwasi P. *Estimation of the Return Periods of Maxima Rainfall and Floods at the Pra River Catchment, Ghana, West Africa using the Gumbel Extreme*

- Value Theory. Heliyon : 7. 2021.
- Gulap S, Gitika T. 2019. *Flood Frequency Analysis Using Gumbel's Distribution Method: A Lower Downstream of Lohit River (Dangori River), Assam (India)*. International Journal of Civil Engineering and Technology(IJCIET). 10(11) : 229-234
- Handiana D, Wahyono S C, Susanti D S. 2013. Perancangan Model Prediksi Curah Hujan Bulanan Berdasarkan Suhu Permukaan Laut di Kalimantan Selatan. Jurnal Fisika Flux. Vol 10 : 1.
- Rini S. 2007. Menentukan metode peramalan yang tepat. Bina Ekonomi Majalah Ilmiah Fakultas Ekonomi Unpar. Vol 11 : 1
- Sasireka K, Suribabu C R, Neelakantan T R. 2019. *Extreme Rainfall Return Preiods using Gumbel and Gamma Distribution*. International Journal of Recent Technology and Engineering (IJRTE). Vol 8 : 482.
- Suripin. 2004. Sistem Drainase Perkotaan yang Berkelanjutan, Yogyakarta: Andi.
- Upomo T C, Kusumawardani W. 2016. Pemilihan Distribusi Probabilitas pada Analisis Hujan dengan Metode Goodness of Fit Test. Jurnal Teknik Sipil dan Perencanaan. 18(2) : 139-148
- Vivekanandan N. 2017. *Distribution for Estimation of Peak Flood Discharge for Ungauged Catchments*. International Journal of Research and Innovation in Social Science (IJRISS). Vol 1 : 6.

Analisis Curah Hujan di Pulau Kalimantan Menggunakan Metode PCA (*Principal Component Analysis*)

Salsabilla Rahmah

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: 2502salsabilla@apps.ipb.ac.id

Abstrak.

Curah hujan merupakan salah satu komponen iklim yang sangat berpengaruh dengan kehidupan manusia. Curah hujan satu wilayah dengan wilayah lain sangat bervariasi karena dipengaruhi oleh faktor lokal maupun global. Dalam skala lokal, posisi geografis dan topografi sangat mempengaruhi variasi curah hujan secara spasial maupun temporal. Salah satunya di pulau Kalimantan yang memiliki curah hujan yang berbeda di setiap provinsinya. Dari 15 kota di 5 provinsi di pulau Kalimantan dari tahun 2001 sampai dengan tahun 2019 merupakan data yang sangat besar untuk dilakukan pengolahan data. Karena besarnya data yang akan diolah maka dapat dilakukan PCA terhadap data tersebut. *Principal Components Analysis* (PCA) atau disebut juga Transformasi Karhunen loeve adalah suatu teknik yang digunakan untuk menyederhanakan suatu data, dengan cara mentransformasi linear sehingga terbentuk sistem koordinat baru dengan variansi maksimum. Hal ini bisa memberikan analisis terhadap data curah hujan tertinggi di 15 kota pada 5 provinsi di Kalimantan dengan data yang jauh lebih sedikit.

Kata Kunci: PCA, Curah hujan, Kalimantan.

1. Pendahuluan

Air merupakan hal yang sangat penting bagi keberlangsungan makhluk hidup di dunia ini. Ketersediaan air merupakan sesuatu yang sangat vital bagi kehidupan umumnya dan manusia khususnya. Berdasarkan dinamika siklus hidrologi salah satu sumber air utama adalah hujan. Ketersediaan air secara alami dalam skala global adalah tetap, hanya terjadi, variasi baik terhadap ruang maupun waktu pada skala regional. Curah hujan sebagai salah satu komponen iklim, sangat penting dalam kehidupan sehari-hari sebagai sumber air yang dibutuhkan makhluk hidup untuk melakukan kegiatan sehari-hari.

Curah hujan satu wilayah dengan wilayah lain sangat bervariasi karena dipengaruhi oleh faktor lokal maupun global. Dalam skala lokal, posisi geografis dan topografi sangat mempengaruhi variasi curah hujan secara spasial maupun temporal. Peningkatan produksi tanaman pangan melalui pemanfaatan data iklim, tidak terlepas dari keberhasilan dalam menginterpretasikan data iklim, terutama curah hujan yang sangat bervariasi. Pada bidang pertanian interpretasi data curah hujan tersebut berkaitan dengan pengambilan keputusan manajemen pertanian, seperti pola tanam dan tata tanam, pengaturan irigasi (Tjasyono, 2004).

Salah satu usaha menginterpretasikan data hujan adalah menyederhanakan data dengan cara menetapkan wilayah-wilayah iklim yang homogen. Pewilayahan hujan dapat diartikan mendeliniasi stasiun curah hujan pada suatu wilayah ke dalam kelompok yang homogen. Homogen dalam bidang hidrologi diartikan sebagai wilayah yang mempunyai kesamaan karakteristik hidrologis, seperti geografis, fisik, statistik atau ciri stokastik (Modarres, 2006).

Data curah hujan merupakan data yang didapatkan dari stasiun-stasiun cuaca yang secara kontinu diukur berulang kali dalam skala hari, minggu atau bulan. Karena diukur secara kontinu dan tidak terputus maka data curah hujan sangat tepat jika dibuat dalam bentuk kurva atau fungsi. Berdasarkan sumber data yang dimiliki yaitu data curah hujan di 15 kota di pulau Kalimantan dari tahun 2001 sampai 2019 maka besarnya dimensi dari data dapat memperbesar kesalahan dalam analisis data maka dari itu akan diterapkan metode *Principal Componen Analysis* (PCA) dengan tujuan untuk mereduksi data sehingga menghasilkan data yang lebih kecil.

Principal Componen Analysis (PCA) adalah sebuah cara untuk mengidentifikasi pola pada data dan kemudian mengekspresikan data tersebut ke bentuk yang lain untuk menunjukkan perbedaan dan persamaan antar pola. Tujuan dari PCA adalah untuk mereduksi dimensi yang besar dari ruang data (*observed variables*) menjadi dimensi yang lebih kecil dari ruang fitur (*independent variables*), yang dibutuhkan untuk mendeskripsikan data lebih sederhana (Lim, 2002).

Menurut Maitra dan Yan (2008), PCA sebagai metodologi dimensi reduksi, diaplikasikan tanpa menganggap adanya korelasi antara variabel bebas dan variabel tak bebas, sedangkan PLS diaplikasikan berdasarkan korelasi. Maka, PCA disebut sebagai *unsupervised dimension reduction methodology*, dan PLS disebut sebagai *supervised dimension reduction methodology*. Meskipun demikian, keduanya memiliki persamaan dalam analisis regresi, yaitu keduanya mempunyai teknik yang digunakan untuk mengubah sebuah set variabel yang berkorelasi tinggi menjadi set dari variabel independen dengan menggunakan transformasi linier dan teknik yang digunakan keduanya adalah untuk mereduksi variabel.

2. Data dan Metode

2.1. Alat dan Bahan

Alat dan bahan yang di gunakan dalam penulisan makalah ini adalah sebagai berikut:

1. *Personal Computer* dengan spesifikasi sebagai berikut :

Spesifikasi	Deskripsi
Nama PC	ACER
<i>System Model</i>	N16Q6
<i>Processor</i>	Intel(R) Celeron(R) Processor N3350
Sistem Operasi	Windows 10 <i>Single Language</i> 64-bit (6.3, <i>Build</i> 9600)
RAM	2.00 GB
Konektivitas	Wi-Fi/Camera/Card Reader/USB/Bluetooth

2. Software yang digunakan sebagai berikut

Nama Software	Kegunaan
Microsoft Excel	untuk membuka data asli dengan format CSV
Julia	untuk mengolah data dengan menggunakan paket yang ada

3. Paket yang di gunakan pada Julia

Nama Paket	Fungsi
CSV.jl	Menangani data teks dengan format delimited, baik itu comma-delimited (csv), tab-delimited (tsv), atau lainnya agar dapat terbaca.
DataFrames.jl	Menyediakan seperangkat <i>tools</i> untuk bekerja dengan data berbentuk tabel di Julia. Desain dan fungsinya mirip dengan pandas (dalam Python) dan data.frame, data.table dan dplyr (dalam R).
Plots.jl	<i>Toolset</i> untuk keperluan visualisasi. Bekerja sama dengan backend lain, seperti GR, PyPlot, PGFPlotsX, atau Plotly.
Measures.jl	Memungkinkan untuk membangun dan menampilkan berbagai jenis unit atau satuan dan kemudian dapat dievaluasi.
MultivariateStats.jl	
LinearAlgebra.jl	Memungkinkan pengoperasian berbasis aritmatika array, faktorisasi matriks dan fungsi terkait aljabar linier dapat dijalankan.
Statistics.jl	
LsqFit.jl	
MLBase.jl	

2.2. Data

Pada percobaan kali ini menggunakan data curah hujan bulanan di 15 kota di Kalimantan yang tersebar di 5 provinsi di Kalimantan. Data 15 kota tersebut antara lain adalah Samarinda, Balikpapan, Sendawar, Tarakan, Tanjung Selor, Malinau, Palangkaraya, Purukcahu, Pangkalbanun, Banjarmasin , Kota Baru, Barabai, Pontianak Ketapang dan Sintang dari tahun 2001 sampai 2019 sumber data adalah sebagai berikut (artikel: baca <https://doi.org/10.1088/1742-6596/1796/1/012053>).

2.3. Metode PCA (*Principal Component Analysis*)

Analisis Komponen Utama (*Principal Component Analysis*) adalah analisis multivariate yang mentransformasi variabel-variabel asal yang saling berkorelasi menjadi variabel-variabel baru yang tidak saling berkorelasi dengan mereduksi sejumlah variabel tersebut sehingga mempunyai dimensi yang lebih kecil namun dapat menerangkan sebagian besar keragaman variabel aslinya.

Banyaknya komponen utama yang terbentuk sama dengan banyaknya variabel asli. Pereduksian (penyederhanaan) dimensi dilakukan dengan kriteria persentase keragaman data yang diterangkan oleh beberapa komponen utama pertama. Apabila beberapa komponen utama pertama telah menerangkan lebih dari 75% keragaman data asli, maka analisis cukup dilakukan sampai dengan komponen utama tersebut.

Bila komponen utama diturunkan dari populasi multivariat normal dengan random vektor $X = (X_1, X_2, \dots, X_p)$ dan vektor rata-rata $\mu = (\mu_1, \mu_2, \dots, \mu_p)$ dan matriks kovarians Σ dengan akar ciri (*eigenvalue*) yaitu $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ didapat kombinasi linier komponen utama yaitu sebagai berikut.

$$Y_1 = e'_1 X = e'_{11} X_1 + e'_{21} X_2 + \cdots + e'_{p1} X_p$$

$$Y_2 = e'_2 X = e'_{12} X_1 + e'_{22} X_2 + \cdots + e'_{p2} X_p$$

⋮

$$Y_p = e'_p X = e'_{1p} X_1 + e'_{2p} X_2 + \cdots + e'_{pp} X_p$$

maka $Var(Y_i) = e'_i \Sigma e_i$ dan $Cov(Y_i, Y_k) = e'_i \Sigma e_k$ dimana $i, k = 1, 2, \dots, p$. Syarat untuk membentuk komponen utama yang merupakan kombinasi linear dari variabel X agar mempunyai varian maksimum adalah dengan memilih vektor ciri (eigen vector) yaitu $e = (e_1, e_2, \dots, e_p)$ sedemikian hingga $Var(Y_i) = e'_i \Sigma e_i$ maksimum dan $e'_i e_i = 1$.

- Komponen utama pertama adalah kombinasi linear $e'_1 X$ yang memaksimumkan $Var(e'_1 X)$ dengan syarat $e'_1 e_1 = 1$.
- Komponen utama kedua adalah kombinasi linear $e'_2 X$ yang memaksimumkan $Var(e'_2 X)$ dengan syarat $e'_2 e_2 = 1$.
- Komponen utama ke- i adalah kombinasi linear $e'_i X$ yang memaksimumkan $Var(e'_i X)$ dengan syarat $e'_i e_k = 0$ untuk $k < i$ dan $Cov(e'_i e_k) = 0$ untuk $k < i$.

Antar komponen utama tersebut tidak berkorelasi dan mempunyai variasi yang sama dengan akar ciri dari Σ . Akar ciri dari matriks ragam peragam Σ merupakan varian dari komponen utama Y, sehingga matriks ragam peragam dari Y adalah:

$$\sum = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}$$

Total keragaman variabel asal akan sama dengan total keragaman yang diterangkan oleh komponen utama yaitu:

$$\sum_{i=1}^p var(X_i) = tr(\sum) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^p var(Y_i)$$

Penyusutan dimensi dari variabel asal dilakukan dengan mengambil sejumlah kecil komponen yang mampu menerangkan bagian terbesar keragaman data. Apabila komponen utama yang diambil sebanyak q komponen, dimana $q < p$, maka proporsi dari keragaman total yang bisa diterangkan oleh komponen utama ke- i adalah:

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}, i = 1, 2, \dots, p$$

Penurunan komponen utama dari matriks korelasi dilakukan apabila data sudah terlebih dahulu ditransformasikan ke dalam bentuk baku Z. Transformasi ini dilakukan terhadap data yang satuan pengamatannya tidak sama. Bila variabel yang diamati ukurannya pada skala dengan perbedaan yang sangat lebar atau satuan ukurannya tidak sama, maka variabel tersebut perlu dibakukan (*standardized*).

Variabel baku (Z) didapat dari transformasi terhadap variabel asal dalam matriks berikut:

$$z = (V^{\frac{1}{2}})^{-1}(X - \mu)$$

$V^{1/2}$ adalah matriks simpangan baku dengan unsur diagonal utama adalah λ_i sedangkan unsur lainnya adalah nol. Nilai harapan $E(Z) = 0$ dan keragamannya adalah

$$Cov(Z) = (V^{\frac{1}{2}})^{-1} \sum (V^{\frac{1}{2}})^{-1} = \rho$$

Dengan demikian komponen utama dari Z dapat ditentukan dari vektor ciri yang didapat melalui matriks korelasi variabel asal ρ . Untuk mencari akar ciri dan menentukan vektor pembobotnya sama seperti pada matriks Σ . Sementara trace matriks korelasi ρ akan sama dengan jumlah p variabel yang dipakai. Pemilihan komponen utama yang digunakan didasarkan pada nilai akar cirinya, yaitu komponen utama akan digunakan jika akar cirinya lebih besar dari satu.

3. Hasil dan Pembahasan

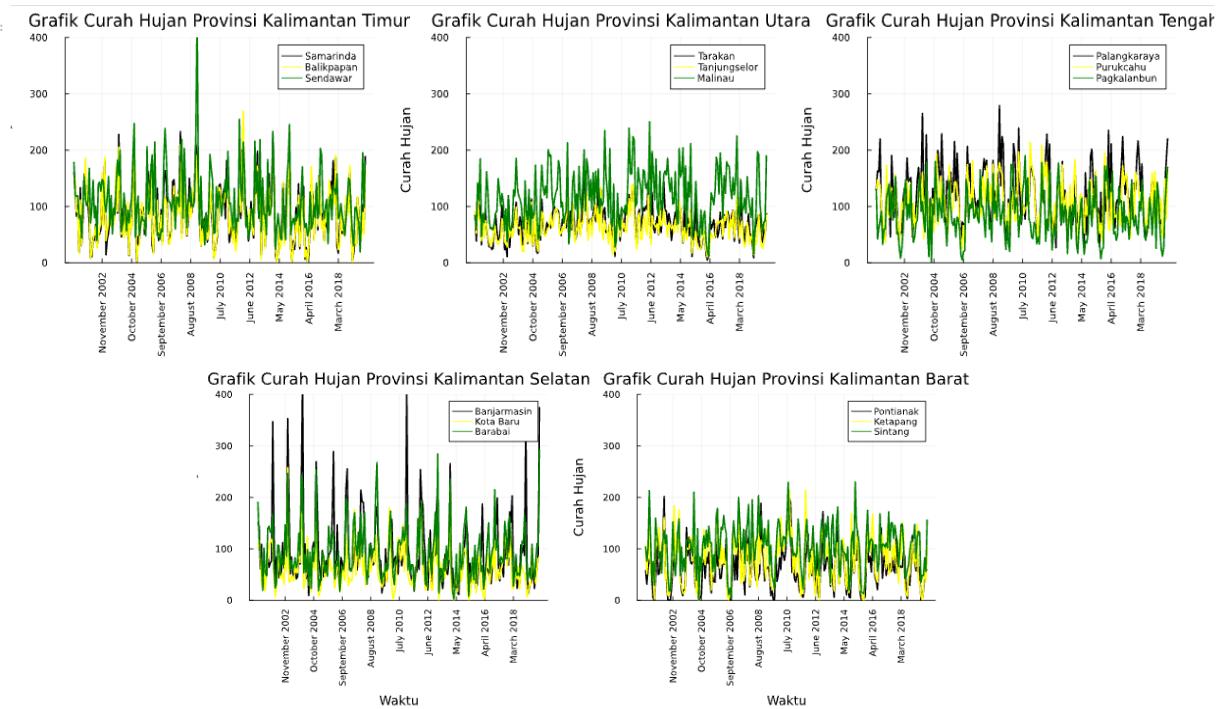
3.1. Deskripsi data

Data yang digunakan adalah data curah hujan 15 kota di pulau Kalimantan yang tersebar di 5 provinsi yaitu Kalimantan Timur, Kalimantan Utara, Kalimantan Tengah, Kalimantan Selatan dan Kalimantan Barat. Data menunjukkan curah hujan berdasarkan waktu mulai dari tahun 2001 hingga tahun 2019. Karena besarnya data maka akan dilakukan reduksi dimensi agar data yang diolah lebih sedikit namun masih menggambarkan keseluruhan data yang ada.

	Waktu	Samarinda	Balikpapan	Sendawar	Tarakan
	String	Float64	Float64	Float64	Float64
1	January 2001	162.79	144.95	179.15	84.99
2	February 2001	133.22	150.68	119.72	38.64
3	March 2001	82.9774	90.6084	118.192	132.794
4	April 2001	101.92	94.18	119.02	83.98
5	May 2001	19.11	16.9	80.44	101.61
6	June 2001	133.49	111.04	83.63	32.17
7	July 2001	30.99	32.01	44.66	81.03
8	August 2001	78.8712	93.6209	100.308	71.4065
9	September 2001	157.07	129.06	152.97	94.74
10	October 2001	151.276	185.813	112.125	61.9074

Gambar 1. Tampilan data menggunakan DataFrame

3.2. Visualisasi Data

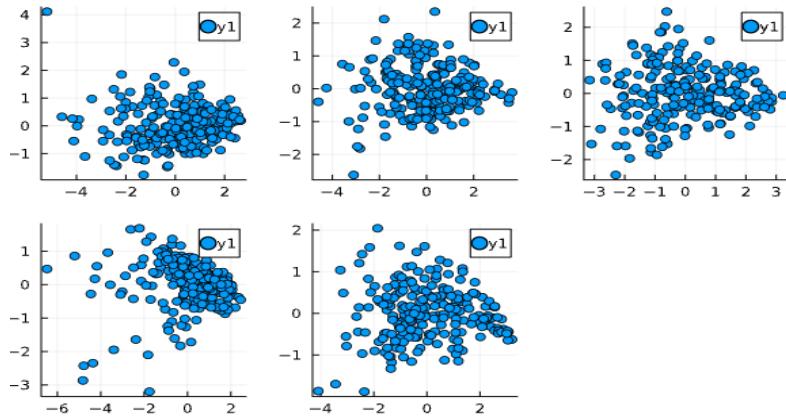


Gambar 2. Visualisasi data menggunakan Plot line

Dari gambar tampak visualisasi data di atas tampak bahwa curah hujan di kota dalam satu provinsi memiliki curah hujan yang serupa.

3.3. Hasil Metode PCA

Dari hasil reduksi dimensi terhadap waktu dengan menggunakan PCA, data di reduksi hingga membentuk 2 dimensi yang menggambarkan lebih dari 80% data yang sebelumnya. Karena dalam proses untuk mereduksi dimensi dari yang semula berdimensi besar menjadi data yang berdimensi relatif kecil maka akan ada sebagian informasi yang hilang namun dengan menggunakan metode PCA dapat meminimalisir informasi yang hilang. Terlihat dari galat dat yang relatif kecil seperti pada data curah hujan di provinsi Kalimantan Barat setelah direduksi menghasilkan galat sebesar 8,3%



Gambar 3. Visualisasi data hasil PCA

4. Simpulan

Kesimpulan yang dapat ditarik dari proses mereduksi dimensi menggunakan metode PCA yaitu pada visual data hasil PCA pada gambar 3 setiap titik pada scatter plot mengandung sejumlah informasi yang menggambarkan data yang lebih besar. Selain itu pada gambar 3 juga tampak sebaran curah hujan pada setiap provinsi

5. Daftar Pustaka

- Tjasyono, Bayong. 2004. *Klimatologi*. Edisi ke-2. ITB.Bandung.
- Goncalves, M.F., Blanco, C.J.C., Santos, V.,C., Oliveira, L.,L.,S., dan Pessoa, F.C.L. 2016. Identification of Rainfall Homogenous Regions Taking into Account El-Nino and LaNina and Rainfall Decrease in the State of Para, Brazilian Amazon. *Acta Scientiarum Technology*. Vol.38. No.2. pp. 209.
- D. Mulyono, 2016. Analisis Karakteristik Curah Hujan di Wilayah Kabupaten Garut Selatan”, Jurnal Konstruksi, vol. 12, no. 1
- Lim, Resmana., Raymond., Kartika. G. 2002. Face Recognition Menggunakan Metode. Linear Discriminant Analysis (LDA). *Proceding Komputer dan Sistem Intelijen*.
- Maitra S, Yan J. 2008. Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression. *Casualty Actuarial Society, 2008 Discussion Paper Program*

Penerapan Metode *K-Means* dan *K-Medoids Clustering* pada Laju Inflasi Beberapa Kategori di Indonesia

Siti Nurhayatudin

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: sitinurhayatudin@apps.ipb.ac.id

Abstrak.

Inflasi merupakan suatu kenaikan harga barang dan jasa secara umum dan terus menerus dalam jangka waktu tertentu. Inflasi berkaitan dengan mekanisme pasar yang dapat disebabkan oleh berbagai faktor, antara lain, konsumsi masyarakat yang meningkat, berlebihnya likuiditas di pasar yang memicu konsumsi atau spekulasi, termasuk juga akibat adanya ketidaklancaran distribusi barang. Inflasi merupakan faktor yang berpengaruh besar terhadap pertumbuhan ekonomi negara Indonesia, akan tetapi banyaknya kategori inflasi cukup menyulitkan dalam proses pengelompokan laju inflasi tersebut. Karya ilmiah ini akan membahas pengelompokan data laju inflasi 3 kategori yaitu transportasi komunikasi jasa keuangan, umum, dan makanan jadi. Data laju inflasi yang digunakan merupakan data pada rentang waktu Januari 2015-Desember 2019. Data laju inflasi akan dibagi menjadi 3 kelompok atau *cluster* yaitu *cluster* dengan inflasi rendah, sedang, dan tinggi. Pengelompokan data laju inflasi dilakukan dengan menggunakan metode *K-Means* dan *K-Medoids clustering*. Tujuan dari karya ilmiah ini adalah menerapkan metode *K-Means* dan *K-Medoids* dalam mengelompokkan laju inflasi, serta membandingkan hasil dari pengelompokan data kedua metode tersebut. Analisis data yang telah dilakukan menunjukkan bahwa pada rentang waktu Januari 2015-Desember 2019, sebagian besar inflasi yang terjadi di Indonesia berdasarkan 3 kategori merupakan inflasi sedang.

Kata Kunci: Inflasi, *Clustering*, *K-Means*, *K-Medoids*.

1. Pendahuluan

Menurut Bank Indonesia inflasi dapat diartikan sebagai kenaikan harga barang dan jasa secara umum dan terus menerus dalam jangka waktu tertentu. Inflasi berkaitan dengan mekanisme pasar yang dapat disebabkan oleh berbagai faktor, antara lain, konsumsi masyarakat yang meningkat, berlebihnya likuiditas di pasar yang memicu konsumsi atau spekulasi, termasuk juga akibat adanya ketidaklancaran distribusi barang. Inflasi timbul karena adanya tekanan dari sisi *supply* (*cost push inflation*), dari sisi permintaan (*demand pull inflation*), dan dari ekspektasi inflasi. Faktor-faktor terjadinya *cost push inflation* dapat disebabkan oleh depresiasi nilai tukar, dampak inflasi luar negeri terutama negara-negara mitra dagang, peningkatan harga-harga komoditi yang diatur pemerintah (*Administered Price*), dan terjadi *negative supply shocks* akibat bencana alam dan terganggunya distribusi. Inflasi yang rendah dan stabil merupakan prasyarat bagi pertumbuhan ekonomi yang berkesinambungan yang pada akhirnya memberikan manfaat bagi peningkatan kesejahteraan masyarakat. Pentingnya pengendalian inflasi didasarkan pada pertimbangan bahwa inflasi yang tinggi dan tidak stabil memberikan dampak negatif kepada kondisi sosial ekonomi masyarakat.

Tingkat inflasi bisa diukur dengan indikator Indeks Harga Konsumen (IHK). Inflasi yang diukur IHK kemudian dikelompokkan ke dalam 7 kelompok pengeluaran yang di antaranya yaitu 1. Kelompok bahan makanan, 2. Kelompok makanan jadi, minuman, rokok dan tembakau, 3. Kelompok perumahan, air, listrik, gas, dan bahan bakar, 4. Kelompok sandang, 5. Kelompok kesehatan, 6. Kelompok pendidikan, rekreasi, dan olahraga, 7. Kelompok transportasi, komunikasi, dan jasa keuangan. Namun tidak semua kelompok pengeluaran akan dianalisis pada karya ilmiah ini, yang akan digunakan hanya 3 kelompok pengeluaran.

Clustering merupakan salah satu metode analisis data yang mengelompokkan *item* data ke dalam sejumlah kecil grup sedemikian sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial (Garcia *et. al* 2002). *Clustering* merupakan salah satu metode yang diterapkan pada *Knowledge in Discovery Database* (Berkhin 2006). *Clustering* terbagi menjadi beberapa metode, salah satunya metode yang sering digunakan yaitu *K-Means* dan *K-Medoids clustering*. Metode *K-Means clustering* cukup efektif untuk diterapkan dalam proses pengklasifikasian karakteristik terhadap objek penelitian. (Darmi dan Setiawan 2016).

Karya ilmiah ini bertujuan untuk mengelompokkan inflasi dari 3 kelompok pengeluaran yaitu inflasi transportasi komunikasi jasa keuangan, inflasi makanan jadi, dan inflasi umum yang terjadi selama bulan Januari 2015-Desember 2019 di Indonesia. Pengelompokan data tersebut dengan dilakukan metode *clustering* yang dibagi menjadi 3 bagian yaitu inflasi tinggi, inflasi sedang, dan inflasi rendah.

2. Metode

2.1. Alat dan Bahan

Proses penggerjaan karya ilmiah ini menggunakan *personal computer* AsusVivoBook 14 Ultra A412FA dengan *processor* Intel Pentium Gold 5405U 2.3GHz dan RAM sebesar 4GB. Karya ilmiah ini juga menggunakan *software Julia* dan *Microsoft Excel* untuk menganalisis data. *Package* yang digunakan untuk menganalisis data pada *software Julia* adalah *Plots*, *XLSX*, *DataFrames*, *StatsBase*, *Statistics*, *LinearAlgebra*, *RDatasets*, *VegaLite*, *VegaDatasets*, *Distances*, dan *Clustering*.

2.2. Data

Sumber data karya ilmiah ini diperoleh dari Kementerian Perdagangan Republik Indonesia dengan situs <https://statistik.kemendag.go.id/inflation>. Data yang digunakan pada karya ilmiah ini merupakan data inflasi dari bulan Januari 2015-Desember 2019. Data inflasi ini terdiri atas tiga kategori yaitu, inflasi transportasi komunikasi dan jasa keuangan, inflasi umum, dan inflasi makanan jadi. Data inflasi nantinya akan dianalisis dengan menggunakan metode *K-Means clustering* berdasarkan kategori inflasi.

2.3. Metode *K-Means Clustering*

Clustering adalah salah satu metode analisis data yang digunakan untuk mengelompokkan data yang berdasarkan kesamaan sifat. Data yang memiliki sifat sama akan dikelompokkan ke dalam satu kelompok atau *cluster*, sedangkan data yang tidak memiliki kesamaan sifat akan dikelompokkan ke dalam kelompok yang lain (Rifa *et. al* 2020). Metode *clustering* bertujuan untuk menimimalkan jarak di dalam *cluster* dan memaksimalkan jarak antar *cluster* (Farahdinna *et. al* 2019). Hal yang penting dalam proses pengelompokan data adalah menyatakan sekumpulan pola ke kelompok yang sesuai untuk menemukan kesamaan dan perbedaan sehingga dapat menghasilkan kesimpulan yang informatif.

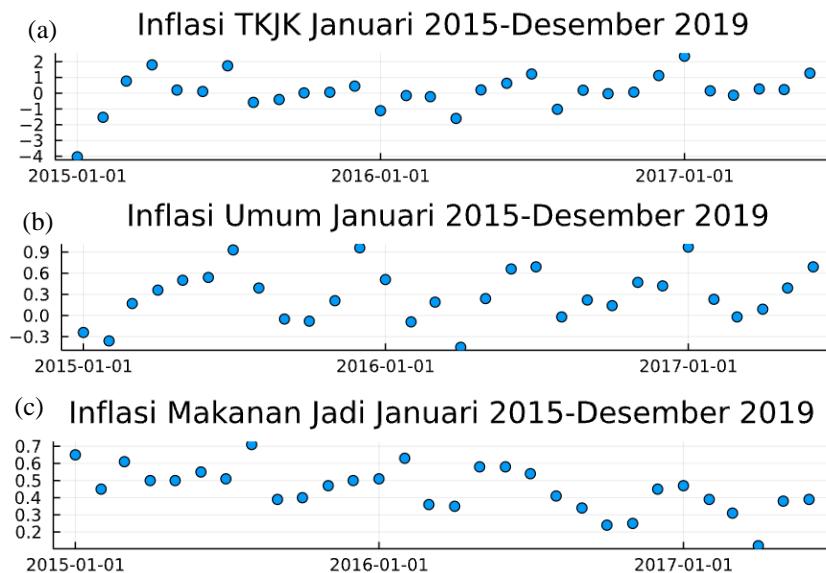
Algoritma *K-Means* merupakan algoritma yang mengelompokkan data berdasarkan titik pusat *cluster* (*centroid*) terdekat dengan data. Tujuan dari *K-Means* adalah mengelompokkan data dengan memaksimalkan kemiripan data dalam satu *cluster* dan meminimalkan kemiripan data antar *cluster* (Sadewo *et. al* 2018). Langkah-langkah *clustering* yang terdapat pada algoritma *K-Means* yaitu 1. Tentukan *k* sebagai jumlah *cluster* yang dibentuk, 2. Tentukan *k centroid* (titik pusat *cluster*) awal secara acak, 3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing cluster, 4. Alokasikan masing-masing objek ke dalam *centroid* yang paling dekat, 5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan, 6. Ulangi langkah 3 jika posisi *centroid* baru tidak sama (Fitriyadi dan Kurniawati 2021).

2.4. Metode *K-Medoids Clustering*

Algoritma *K-Medoids* adalah algoritma *clustering* yang mirip dengan *K-Means*. Perbedaan dari kedua algoritma ini yaitu algoritma *K-Medoids* menggunakan objek sebagai perwakilan (*medoid*) sebagai pusat *cluster* untuk setiap *cluster* (Surya dan Laurence 2019). Langkah-langkah algoritma *K-Medoids* yaitu 1. Inisialisasi pusat *cluster* sebanyak *k* (jumlah *cluster*), 2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian*, 3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoid* baru, 4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoid* baru, 5. Hitung total simpangan (*S*) dengan menghitung nilai total jarak baru-total jarak lama, 6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoid*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing (Fitriyadi dan Kurniawati 2021).

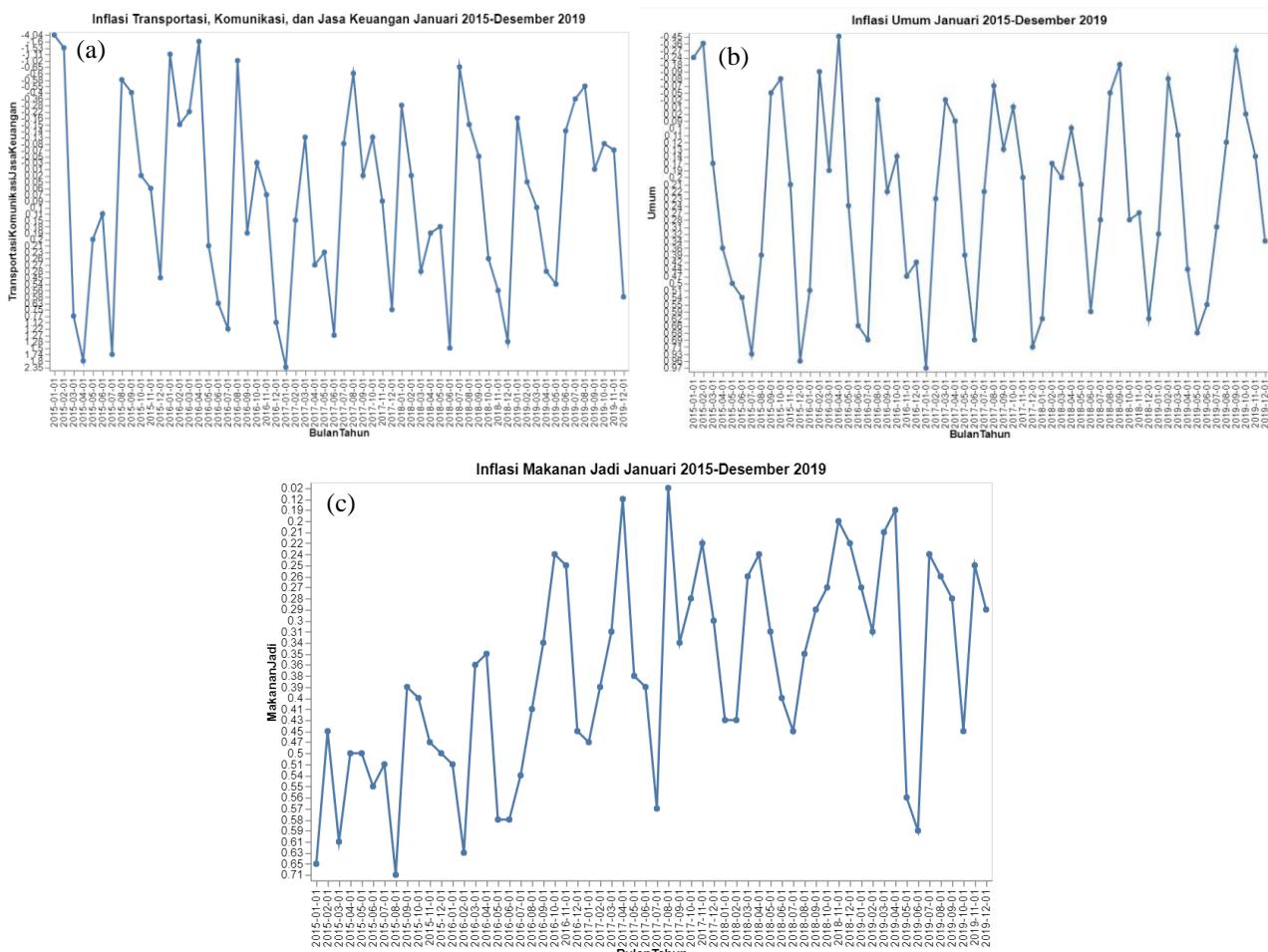
3. Hasil dan Pembahasan

Data laju inflasi yang digunakan pada karya ilmiah ini akan dianalisis dengan menggunakan metode *K-Means Clustering* dan *K-Medoids Clustering*. Langkah pertama yang dilakukan dalam analisis data adalah membuat *scatter plot* data dengan *software Julia* yang dilakukan pada *Jupyter Notebook*. Kemudian dapat terlihat dari *scatter plot* bahwa data laju inflasi ketiga kategori di Indonesia pada bulan Januari 2015-Desember 2019 cenderung fluktuatif.



Gambar 1. Scatter plot laju inflasi bulan Januari 2015-Desember 2019 di Indonesia (a) Kategori transportasi, komunikasi, dan jasa keuangan (b) Kategori umum (c) Kategori makanan jadi.

Plot data juga dilakukan dengan menggunakan *Package VegaLite* dari *software Julia*. Plot laju inflasi yang ditampilkan dengan *VegaLite plot* juga menunjukkan kecenderungan yang fluktuatif pada data. Namun *VegaLite plot* lebih informatif, jelas, dan lengkap dalam menunjukkan informasi data dibandingkan dengan *scatter plot*. Kelemahan dari *VegaLite Plot* adalah laju inflasi pada sumbu y tidak dapat diurutkan berdasarkan nilai laju inflasi terkecil sampai terbesar.



Gambar 2. VegaLite plot laju inflasi bulan Januari 2015-Desember 2019 di Indonesia (a) Kategori transportasi, komunikasi, dan jasa keuangan (b) Kategori umum (c) Kategori makanan jadi.

Pengolahan data laju inflasi pada bulan Januari 2015-Desember 2016 dilakukan dengan menggunakan *software Julia*. *Package* yang diperlukan untuk memperoleh *cluster* pada data adalah *Package Clustering*. Analisis data yang pertama dilakukan adalah analisis dengan menggunakan *K-Means Clustering*. Kelompok atau *cluster* pada data laju inflasi dengan metode *K-Means* diperoleh dengan menjalankan kode 1, kemudian cluster tersebut ditampilkan pada tabel 1 dengan nama kolom *cluster*.

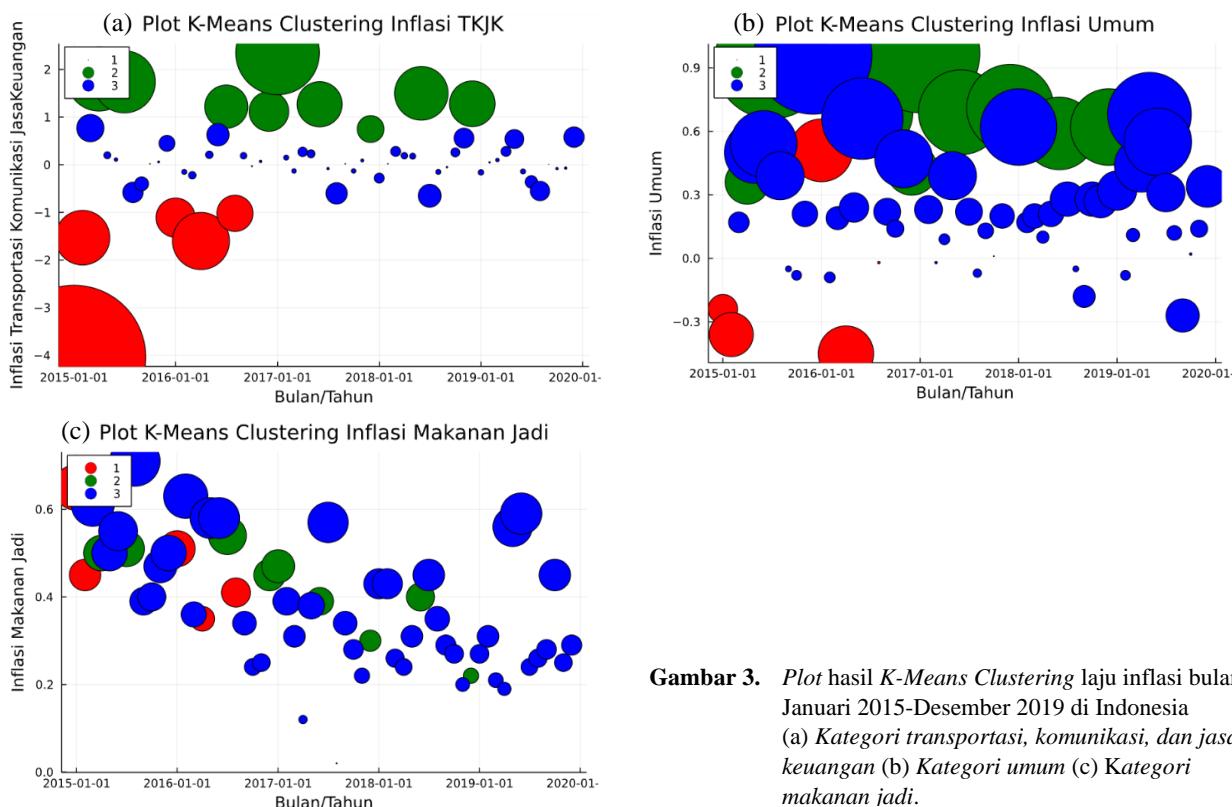
Kode 1.

```
Q1 = df[!, [:TransportasiKomunikasiJasaKeuangan, :Umum, :MakananJadi]]
C = kmeans(Matrix(Q1)', 3)
insertcols!(df, 3,:cluster=>C.assignments)
```

Tabel 1. Hasil *K-Means Clustering* Laju Inflasi Bulan Januari 2015-Desember 2019 di Indonesia

Bulan/Tahun	Cluster	TKJK	Laju Inflasi	Umum	Makanan Jadi
2015-01-01	1	-4.04		-0.24	0.65
2015-02-01	1	-1.53		-0.36	0.45
2015-03-01	3	0.77		0.17	0.61
2015-04-01	2	1.8		0.36	0.5
2015-05-01	3	0.2		0.5	0.5
2015-06-01	3	0.11		0.54	0.55
2015-07-01	2	1.74		0.93	0.51
2015-08-01	3	-0.58		0.39	0.71
2015-09-01	3	-0.4		-0.05	0.39
2015-10-01	3	0.02		-0.08	0.4
:	:	:		:	:

Berdasarkan kelompok atau *cluster* masing-masing data laju inflasi pada algoritma *K-Means*, *cluster* 1 menunjukkan laju inflasi rendah, *cluster* 2 menunjukkan laju inflasi tinggi, dan *cluster* 3 menunjukkan laju inflasi sedang. *cluster* 1 atau laju inflasi rendah memiliki 5 anggota yang masuk dalam kategori, yaitu bulan Januari 2015, Februari 2015, Januari 2016, April 2016, dan Agustus 2016. Selanjutnya, terdapat 9 data yang masuk dalam kategori *cluster* 2 atau laju inflasi tinggi yaitu April 2015, Juli 2015, Juli 2016, Desember 2016, Januari 2017, Juni 2017, Desember 2017, Juni 2018, dan Desember 2018. Data yang masuk dalam kategori *cluster* 3 atau inflasi tinggi ada sebanyak 46 data (bulan/tahun) dalam rentang waktu 2015-2019.



Gambar 3. Plot hasil *K-Means Clustering* laju inflasi bulan Januari 2015-Desember 2019 di Indonesia
(a) Kategori transportasi, komunikasi, dan jasa keuangan (b) Kategori umum (c) Kategori makanan jadi.

Gambar 3 merepresentasikan bahwa setiap kategori inflasi hasil dari algoritma *K-Means* memiliki jumlah anggota *cluster* paling banyak pada cluster ke 3 atau *cluster* dengan inflasi sedang. Kategori inflasi yang anggotanya cenderung menyebar yaitu inflasi umum, dan makanan jadi. Kecenderungan inflasi rendah terjadi pada awal tahun seperti bulan januari, sedangkan kecenderungan inflasi tinggi itu terjadi pada akhir seperti bulan Desember.

Selanjutnya, analisis data yang dilakukan adalah analisis dengan menggunakan *K-Medoids Clustering*. Kelompok atau *cluster* pada data laju inflasi dengan metode *K-Means* diperoleh dengan menjalankan kode 2, kemudian *cluster* tersebut ditampilkan pada tabel 2 dengan nama kolom *cluster1*.

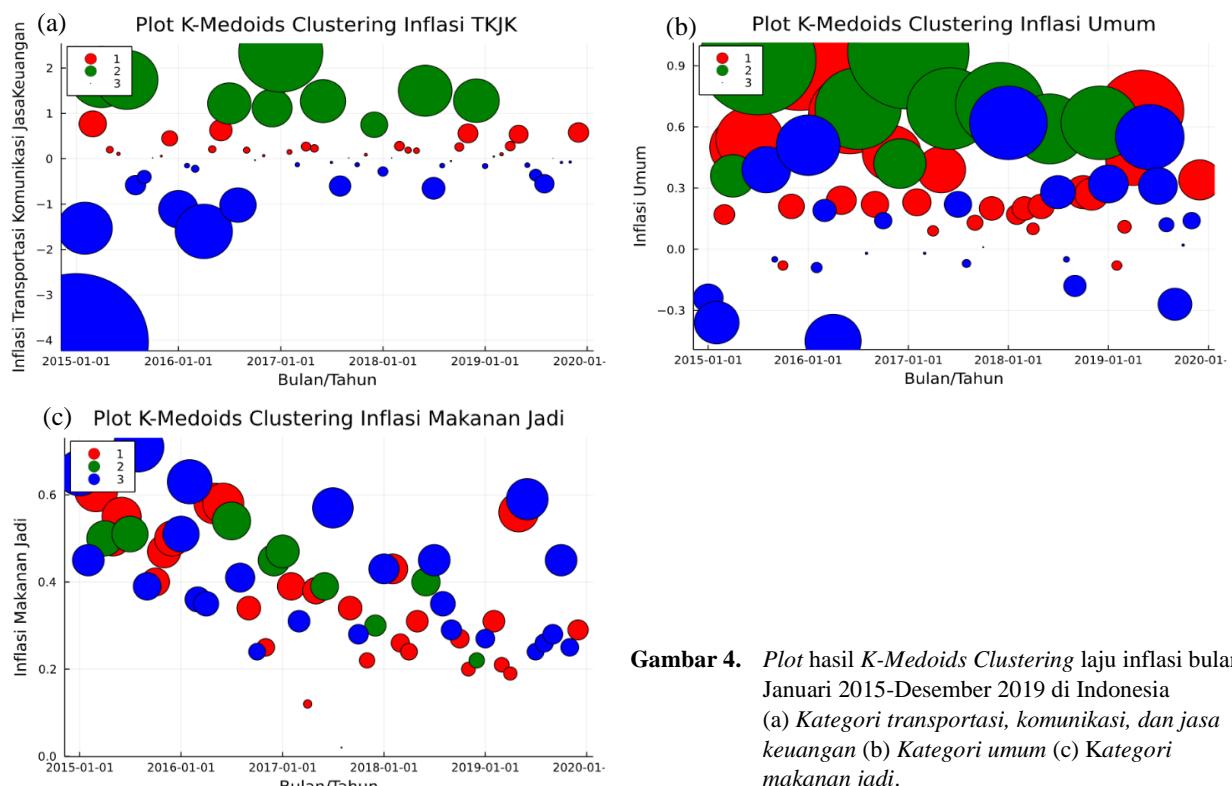
Kode 2.

```
Q1 = df[!, [:TransportasiKomunikasiJasaKeuangan, :Umum, :MakananJadi]]
D = pairwise(Euclidean(), Matrix(Q1)', Matrix(Q1)', dims=2)
K = kmedoids(D, 3)
insertcols!(df, 3, :cluster1=>K.assignments)
```

Tabel 2. Hasil *K-Medoids Clustering* Laju Inflasi Bulan Januari 2015-Desember 2019 di Indonesia

Bulan/Tahun	Cluster1	Cluster	Laju Inflasi		
			TKJK	Umum	Makanan Jadi
2015-01-01	3	1	-4.04	-0.24	0.65
2015-02-01	3	1	-1.53	-0.36	0.45
2015-03-01	1	3	0.77	0.17	0.61
2015-04-01	2	2	1.8	0.36	0.5
2015-05-01	1	3	0.2	0.5	0.5
2015-06-01	1	3	0.11	0.54	0.55
2015-07-01	2	2	1.74	0.93	0.51
:	:	:	:	:	:

Berdasarkan kelompok atau *cluster* masing-masing data laju inflasi pada algoritma *K-Medoids*, *cluster* 1 menunjukkan laju inflasi sedang, *cluster* 2 menunjukkan laju inflasi tinggi, dan *cluster* 3 menunjukkan laju inflasi sedang. *cluster* 1 atau laju inflasi sedang memiliki 26 anggota yang masuk dalam kategori. Selanjutnya, terdapat 9 data yang masuk dalam kategori *cluster* 2 atau laju inflasi tinggi yaitu April 2015, Juli 2015, Juli 2016, Desember 2016, Januari 2017, Juni 2017, Desember 2017, Juni 2018, dan Desember 2018. Data yang masuk dalam kategori *cluster* 3 atau inflasi tinggi ada sebanyak 25 data (bulan/tahun) dalam rentang waktu 2015-2019. Hal yang berbeda antara algoritma *K-Means* dan *K-Medoids* adalah pengelompokan anggota dalam *cluster* sedang dan rendah, karena pada *K-Means* anggota *cluster* dengan laju inflasi rendah hanya berjumlah 5 orang sedangkan pada *K-Medoids* jumlah anggota *cluster* dengan laju inflasi rendah dan sedang berimbang. Hal yang unik pada kedua algoritma yang dilakukan adalah pengelompokan *cluster* 2 atau *cluster* dengan laju inflasi tinggi memiliki jumlah anggota yang tepat sama.



Gambar 4. Plot hasil *K-Medoids Clustering* laju inflasi bulan Januari 2015-Desember 2019 di Indonesia
(a) Kategori transportasi, komunikasi, dan jasa keuangan (b) Kategori umum (c) Kategori makanan jadi.

Gambar 4 merepresentasikan bahwa setiap kategori inflasi hasil dari algoritma *K-Means* memiliki jumlah anggota *cluster* paling banyak pada cluster ke 1 atau *cluster* dengan inflasi sedang dan *cluster* ke 3 atau *cluster* dengan inflasi rendah. Kategori inflasi yang anggotanya cenderung menyebar yaitu inflasi umum, dan makanan jadi. Kecenderungan inflasi rendah terjadi pada awal tahun seperti bulan januari, sedangkan kecenderungan inflasi tinggi itu terjadi pada akhir seperti bulan Desember.

4. Simpulan

Hasil Pengelompokan data laju inflasi dengan menggunakan algoritma *K-Means* dan *K-Medoids* mudah dan cukup cepat untuk digunakan oleh data yang akan dianalisis. Baik pada *K-Means* dan *K-Medoids* mengelompokkan laju inflasi yang tinggi dalam kelompok yang sama, sedangkan untuk *cluster* laju inflasi rendah pada *K-Means* memiliki anggota yang lebih sedikit dibandingkan pada *K-Medoids*. Kedua metode tersebut sudah cukup baik dalam mengelompokkan data, tetapi diperlukan analisis lebih lanjut untuk mengetahui metode manakah yang lebih baik untuk menganalisis data laju inflasi ini.

5. Daftar Pustaka

- Berkhin P. 2006. *A Survey Of Clustering Data Mining Techniques* In: *Grouping multidimensional data*. Jerman(DE): Springer Berlin Heidelberg.
- BI. 2020. Inflasi. [diakses 2021 Okt 09]. <https://www.bi.go.id/id/fungsi-utama/moneter/inflasi/default.aspx>.
- Farahdinna F, Nurdiansyah I, Suryani A, Wibowo A. 2019. Perbandingan algoritma k-means dan k-medoids dalam klasterisasi produk asuransi perusahaan nasional. *J. Ilm. FIFO*. 11(2): 208. doi: 10.22441/fifo.2019.v11i2.010.
- Fitriyadi AU, Kurniawati A. 2021 Analisis algoritma k-means dan k-medoids untuk clustering data kinerja karyawan pada perusahaan perumahan nasional. *Kilat*. 10(1): 2655-4925. doi: 10.33322/kilat.v10i1.1174
- Garcia, HM, Ullman J, Widom, Jennifer. 2002. *Database systems: The complete book*.
- I. H. Rifa, H. Pratiwi, and R. Respatiwulan. 2020. Clustering of earthquake risk in indonesia using k-medoids and k-means algorithms. *Media Stat.* 13(2): 194–205, 2020, doi: 10.14710/medstat.13.2.194-205.
- M. G. Sadewo et al. 2018. Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa / Kelurahan Menurut Upaya Antisipasi / Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means. (2): 311–319.
- Surya P, Laurence AI. 2019. Performance analysis of k-means and k-medoid clustering algorithms using agriculture dataset. *J. Emerg. Technol. Innov. Res.* 6(1):539–545.
- Darmi Y, Setiawan A. 2016. Penerapan metode clustering k-means dalam pengelompokan penjualan produk. *J. Media Infotama*, 12(2): 148–157.

Meramalkan Perubahan Inflasi Pada Tahun 2020 Setiap Bulannya

Rahmawati Masithoh

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: r27masithohrahmawati@apps.ipb.ac.id

Abstrak.

Inflasi merupakan salah satu indikator stabilitas perekonomian. Jika tingkat inflasi rendah dan stabil akan menjadi stimulator pertumbuhan ekonomi. Setiap kali ada gejolak sosial, politik, dan ekonomi di dalam maupun di luar negeri masyarakat selalu mengaitkan dengan masalah inflasi. Pemerintah perlu membuat rencana kebijakan dalam menghadapi perubahan inflasi. Oleh karena itu, pemerintah membutuhkan data tentang perubahan inflasi di masa mendatang, yaitu data hasil peramalan. Sebelum meramalkan, dilihat terlebih dahulu pola data inflasi. Data bulanan inflasi umum dari tahun 2000 sampai 2019 memiliki pola data musiman bulanan dengan mengulangnya fluktuasi data setiap enam bulan sekali yang berarti data ini memiliki panjang musiman $L = 6$. Pola yang lebih mendekati model data ini adalah data musiman. Selain itu, besarnya efek musiman pada data tersebut tetap dari waktu ke waktu. Untuk itu metode yang sesuai adalah Metode Winters (*triple exponential smoothing*) dengan tipe *additive*. Menggunakan parameter pemulusan keseluruhan $\alpha = 0,515$, tren $\gamma = 0,035$, dan musiman $\beta = 0,0825$ menghasilkan data ramalan inflasi umum di tahun 2020 diperoleh bulan Januari besar inflasi 0.35, bulan Februari -0.08, bulan Maret -0.19, bulan April -0.13, bulan Mei 0.04, bulan Juni 0.34, bulan Juli 0.31, bulan Agustus -0.12, bulan September -0.23, bulan Oktober -0.17, bulan November -0.0, dan bulan Desember 0.29 dengan besar kesalahan MAPE adalah 269.99% yang artinya peramalan tidak akurat.

Kata Kunci: Inflasi, *Forecasting*, Holt-Winter.

1. Pendahuluan

Inflasi merupakan kenaikan harga-harga umum yang berlaku dalam suatu perekonomian dari suatu periode ke periode lainnya. Inflasi merupakan salah satu indikator stabilitas perekonomian. Jika tingkat inflasi rendah dan stabil akan menjadi stimulator pertumbuhan ekonomi. Setiap kali ada gejolak sosial, politik, dan ekonomi di dalam maupun di luar negeri masyarakat selalu mengaitkan dengan masalah inflasi (Mankiw, 2006).

Inflasi di Indonesia juga dipengaruhi oleh kenaikan harga komoditi impor (*imported inflation*) dan membengkaknya hutang luar negeri akibat dari terdepresiasinya nilai tukar rupiah terhadap dolar Amerika dan mata uang asing lainnya (Atmadja, 1999). Dengan semakin meningkatnya hutang luar negeri yang dilakukan oleh pemerintah, maka peningkatan defisit anggaran dapat mempengaruhi variabel ekonomi makro seperti tingkat inflasi, tingkat pengangguran, tingkat bunga, jumlah uang beredar, serta variabel lainnya. Defisit anggaran mempengaruhi tingkat inflasi baik itu yang dibiayai dengan penciptaan uang maupun dengan hutang. Defisit ini mempengaruhi inflasi melalui jumlah uang beredar. Semakin besar defisit maka jumlah uang beredar juga semakin besar yang akibatnya inflasi juga membesar. Perhitungan inflasi dilakukan melalui dua pendekatan yakni Indeks Harga Konsumen dan Indeks Harga Produsen (IHP). Indeks Harga Konsumen yang dikenal sebagai IHK atau CPI (*Consumer Price Index*) yang mengukur biaya dari pasar konsumsi barang dan jasa (Hill, 1996).

Salah satu upaya dari pemerintah guna mengatasi masalah inflasi ini adalah dengan melakukan kebijakan fiskal dan kebijakan moneter. Kebijakan fiskal dengan cara pengurangan pengeluaran pemerintah yang dilaksanakan oleh kementerian keuangan sedangkan kebijakan moneter dilakukan oleh Bank Indonesia (Triwahyuni, 2021). Oleh karena itu, pemerintah membutuhkan data tentang perubahan inflasi di masa mendatang, yaitu dengan menggunakan data hasil peramalan untuk merencanakan kebijakan tersebut. Hal ini yang mendasari penulis untuk membuat makalah ini.

2. Data dan Metode

Proses peramalan tentang perubahan inflasi pada tahun 2020 setiap bulannya, menggunakan data dan metode sebagai berikut.

2.1. Alat dan Bahan

Alat dan bahan yang digunakan adalah *personal computer* (PC) dengan jenis sistem operasi Windows 10 Home Single Language 64-bit, jenis *processor* AMD Ryzen 3 2200U with Radeon Vega Mobile Gfx, dan *memory* 4096MB RAM. Untuk melakukan peramalan menggunakan alat bantu *software* Julia tipe 1.6.2.

2.2. Data

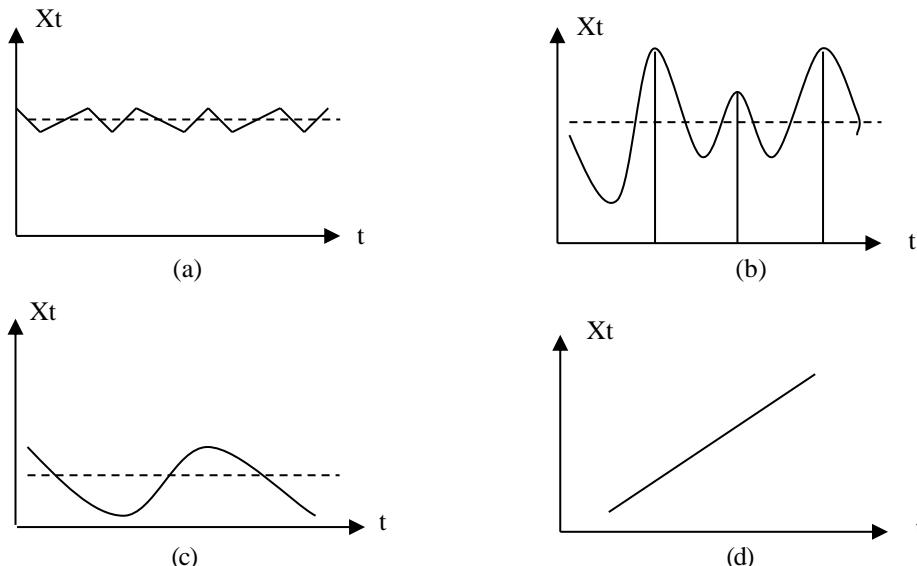
Data yang digunakan merupakan data inflasi jenis umum Indeks Harga Konsumen (IHK) dan berupa data bulanan dari tahun 2000 sampai tahun 2019.

2.3. Metode

Peramalan merupakan bagian integral dari kegiatan pengambilan keputusan, sebab efektif atau tidaknya suatu keputusan umumnya bergantung pada beberapa faktor yang tidak dapat dilihat pada waktu keputusan itu diambil (Aswi dan Sukarna, 2006). Teknik peramalan dibagi ke dalam dua kategori utama, yaitu metode kuantitatif dan metode kualitatif. Metode peramalan kuantitatif dibagi menjadi dua jenis model peramalan utama, yaitu model deret berkala (*time series*) dan model regresi (kausal). Pada model *time series*, pendugaan masa depan dilakukan berdasarkan nilai masa lalu dari suatu variabel dan/atau kesalahan masa lalu. Tujuan metode peramalan deret berkala seperti itu adalah menemukan pola dalam deret data historis dan mengekstrapolasikan pola dalam deret data historis dan mengekstrapolasikan pola tersebut ke masa depan (Makridakis et al, 1999).

Pola data dapat dibedakan menjadi empat, yaitu :

1. Pola horizontal (gambar 1 (a)), terjadi bilamana data berfluktuasi di sekitar nilai rata-rata yang konstan atau stasioner terhadap nilai rata-ratanya.
2. Pola musiman (gambar 1 (b)), terjadi bilamana suatu deret data dipengaruhi oleh faktor musiman (misalnya kuartal tahun tertentu, bulanan atau hari pada minggu tertentu).
3. Pola siklis (gambar 1 (c)), terjadi bilamana datanya dipengaruhi oleh fluktuasi ekonomi jangka panjang seperti yang berhubungan dengan siklus bisnis atau ekonomi.
4. Pola tren (gambar 1 (d)), terjadi bilamana terdapat kenaikan atau penurunan jangka panjang dalam data.



Gambar 1 Macam-macam pola data

Exponential smoothing adalah suatu metode peramalan rata-rata bergerak yang melakukan pembobotan menurun secara *exponential* terhadap nilai observasi yang lebih tua. Metode *exponential smoothing* merupakan pengembangan dari metode *moving average* (Makridakis et al, 1999). Terdapat beberapa metode *Exponential Smoothing*, yaitu Metode Single *Exponential Smoothing*, Metode double *Exponential Smoothing*, dan Metode triple *Exponential Smoothing*. Metode Single *Exponential Smoothing* meramalkan untuk periode sebelumnya dan menyesuaikan dengan menggunakan kesalahan peramalan (Makridakis et al, 1997).

Metode *Holt* merupakan salah satu metode *Double Exponential Smoothing*. Holt memperluas metode *single Exponential Smoothing* menjadi persamaan linier pemulusan eksponensial untuk meramalkan data yang mengandung *trend*. Metode Holt telah diperluas oleh Winters pada tahun 1960 untuk menggambarkan musiman secara langsung. Metode Holt-Winters didasarkan pada tiga persamaan pemulusan, yaitu satu untuk unsur tingkat, satu untuk unsur trend, dan satu untuk unsur musiman. Berdasarkan pada besar efek musimannya model Winters terbagi menjadi dua, yaitu *multiplicative seasonality* dan *additive seasonality* (Makridakis et al, 1997).

a. Multiplicative Seasonality

Bila besarnya efek musiman berubah dari waktu ke waktu, maka bentuk model yang dipakai adalah *multiplicative seasonality*. Persamaan dasar untuk metode *Holt-Winters* dengan *multiplicative seasonality* adalah sebagai berikut:

$$S_t = \alpha \frac{x_t}{l_{t-L}} + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad (1)$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1} \quad (2)$$

$$I_t = \beta \frac{X_t}{S_t} + (1 - \beta) I_{t-L}, \quad (3)$$

$$F_{t+m} = (S_t + b_t m) I_{t-L+m} \quad (4)$$

Pada persamaan (1) L adalah panjang musiman dan S_t merepresentasikan nilai pemulusan keseluruhan periode t . Pada persamaan (2) b_t dinotasikan sebagai nilai pemulusan trend pada periode t . Pada persamaan (3) I_t adalah komponen musiman. Persamaan (3) membandingkan indeks musiman yang merupakan perbandingan observasi X_t dengan nilai pemulusan keseluruhan S_t . Akhirnya didapatkan persamaan (4) F_{t+m} merupakan peramalan untuk m periode mendatang.

Dibutuhkan nilai awal atau inisialisasi dari masing-masing komponen untuk memulai perhitungan. Untuk peramalan metode *Holt Winters* dengan *multiplicative seasonality* dibutuhkan nilai awal untuk pemulusan keseluruhan S_L , pemulusan trend b_L , dan pemulusan musiman I_L . Untuk menghitung estimasi nilai awal dari indeks musiman dibutuhkan setidaknya data lengkap selama satu musim. Oleh karena itu, nilai tren dan nilai pemulusan diinisialisasi pada periode L . Pada pemulusan keseluruhan diinisialisasi dengan mengambil rata-rata pada musim pertama dengan persamaan sebagai berikut

$$S_L = \frac{1}{L} (X_1 + X_2 + \dots + X_L). \quad (5)$$

Persamaan (5) merupakan rata-rata bergerak berorde L yang akan mengeliminasi unsur musimana pada data. Untuk menginisialisasi trend, lebih baik menggunakan data lengkap secara dua musim, yaitu sebagai berikut

$$b_L = \frac{1}{L} \left(\frac{X_{L+1}-X_1}{L} + \frac{X_{L+2}-X_2}{L} + \dots + \frac{X_{L+L}-X_L}{L} \right). \quad (6)$$

Kemudian didapatkan nilai inisialisasi indeks musiman dengan menggunakan rasio dari data tahun pertama dengan rata-rata data satu musim pertama, sehingga

$$I_1 = \frac{X_1}{S_L}, I_2 = \frac{X_2}{S_L}, \dots, I_L = \frac{X_L}{S_L}. \quad (7)$$

Parameter-parameter α , γ , dan β yang bernilai antara 0 dan 1 dapat dipilih untuk meminimalkan nilai error atau kesalahan. Pendekatan untuk menentukan nilai ini biasanya menggunakan cara coba dan salah (*trial and error*).

b. Additive seasonality

Komponen musiman dalam metode *Holt Winters* dapat bersifat aditif bila besarnya efek musiman konstan dari waktu ke waktu, meskipun hal ini jarang terjadi. Persamaan dasar untuk metode *Holt Winters* dengan *additive seasonality* adalah sebagai berikut:

$$S_t = \alpha(X_t - I_{t-L}) + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad (8)$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1} \quad (9)$$

$$I_t = \beta(X_t - S_t) + (1 - \beta)I_{t-L} \quad (10)$$

$$F_{t+m} = S_t + b_t m + I_{t-L+m} \quad (11)$$

Persamaan (8) identik dengan persamaan (2). Perbedaan antara persamaan *additive seasonality* dengan persamaan *multiplicative seasonality* adalah indeks musiman pada persamaan ini dijumlahkan dan dikurangi daripada persamaan sebelumnya yang dikali dan dibagi.

Untuk menginisialisasi nilai awal untuk pemulusan keseluruhan S_L dan pemulusan trend b_L adalah sama dengan metode *multiplicative seasonality*. Sedangkan untuk menginisialisasi nilai awal indeks musiman menggunakan

$$I_1 = X_1 - S_L, I_2 = X_2 - S_L, \dots, I_L = X_L - S_L \quad (12)$$

Mean Absolute Percentage Error (MAPE) merupakan salah satu ukuran ketepatan metode peramalan yang memberikan petunjuk seberapa besar persentase nilai absolut kesalahan peramalan $|X_t - F_t|$ dibandingkan dengan nilai sebenarnya X_t yang kemudian dirata-ratakan. Semakin kecil nilai MAPE, maka akan semakin tepat hasil peramalan yang diperoleh. Persamaan yang digunakan untuk memperoleh nilai MAPE pada suatu hasil peramalan adalah:

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{X_t - F_t}{X_t} \right|}{n} \times 100\% \quad (13)$$

Menurut Goh dan Law (2002), klasifikasi nilai MAPE pada peramalan adalah sebagai berikut.

Tabel 1 Klasifikasi nilai MAPE

Nilai MAPE	Keterangan
< 10%	Sangat akurat
10% - 20%	Baik
20% - 50%	Layak
>50%	Tidak akurat

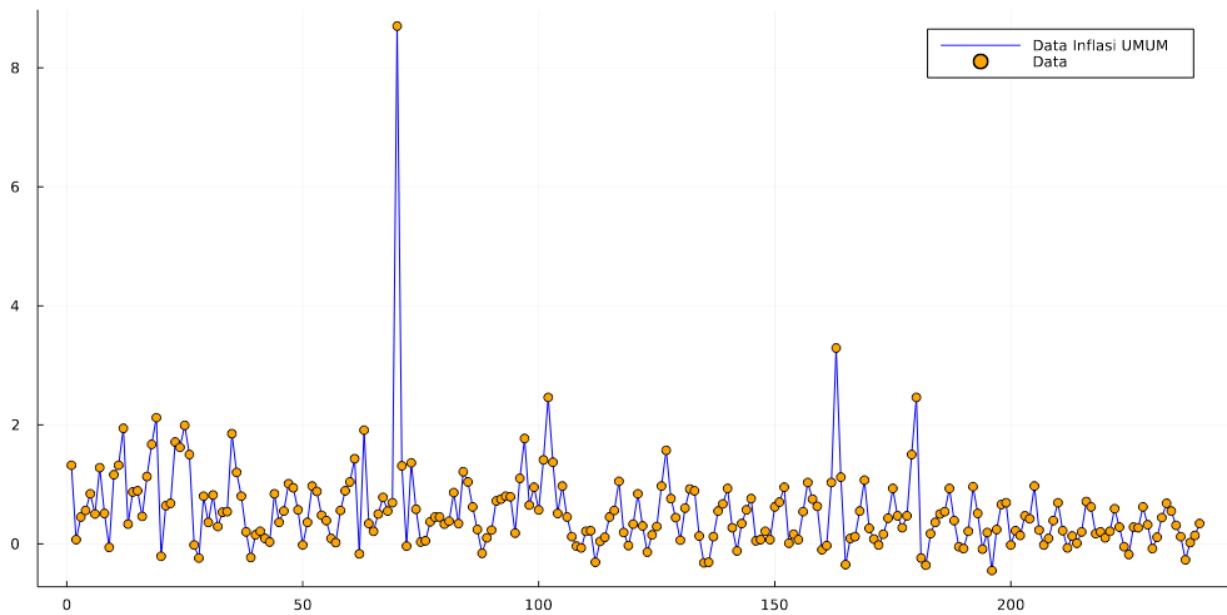
2.5. Tahapan

Peramalan yang baik adalah peramalan yang dilakukan dengan mengikuti prosedur penyusunan yang baik. Berikut tahapan melaksanakan peramalan.

1. Membuat plot data masa lalu dan menentukan metode yang cocok sesuai pola pada plot data.
2. Menentukan nilai awal dari pemulusan-pemulusan pada metode *Holt Winters*.
3. Menghitung setiap parameter fungsi ramalan untuk masing-masing metode *Holt Winters*.
4. Melakukan peramalan data inflasi umum dari tahun 2000-2019.
5. Melakukan perhitungan untuk nilai kesalahan ramalan dengan menggunakan hasil peramalan dan data masa lalu.
6. Memilih parameter fungsi terbaik yang memberikan kesalahan ramalan terkecil untuk metode *Holt Winters*.
7. Melakukan perhitungan untuk nilai kesalahan ramalan dengan parameter pemulusan yang optimal.
8. Melakukan peramalan 12 periode ke depan, yaitu bulan Januari sampai Desember di tahun 2020.

3. Hasil dan Pembahasan

Tahap pertama dalam melakukan peramalan adalah membuat plot data masa lalu, yaitu data bulanan inflasi jenis umum pada tahun 2000 sampai 2019. Berikut hasil plot data tersebut.



Gambar 2. Plot data bulanan inflasi umum tahun 2000-2019

Berdasarkan tabel di atas, data inflasi umum cenderung memiliki pola data musiman bulanan dengan mengulangnya fluktuasi data setiap enam bulan sekali yang berarti data ini memiliki panjang musiman $L = 6$. Menurut Makridakis dkk (1999:11) bahwa pola yang lebih mendekati model data ini adalah data musiman. Perhatikan pula bahwa besarnya efek musiman tetap dari waktu ke waktu. Untuk itu metode yang sesuai adalah metode Winters (*triple exponential smoothing*) dengan tipe *addictive*.

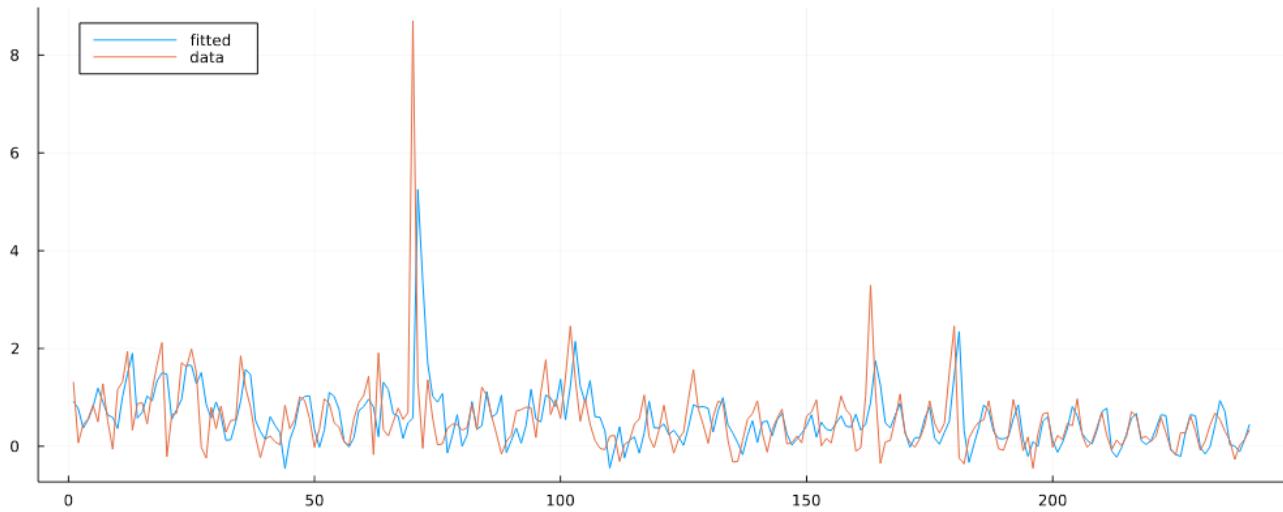
Tahap kedua menentukan nilai awal dari pemulusan-pemulusan pada metode *Holt Winters*, yaitu sebagai berikut.

Nilai awal pemulusan keseluruhan = 0.62.

Nilai awal pemulusan tren = 0.07.

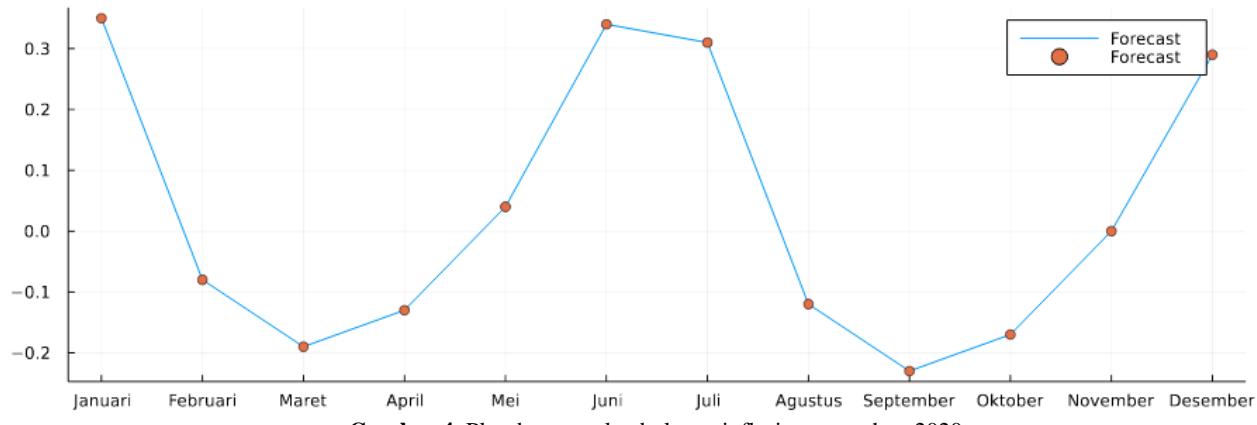
Nilai awal pemulusan musiman = [0.7, -0.55, -0.17, -0.06, 0.22, -0.12]

Selanjutnya, menghitung setiap parameter fungsi ramalan dan melakukan peramalan dengan menggunakan parameter $\alpha = 0,515$, $\gamma = 0,035$, dan $\beta = 0,0825$ menghasilkan data ramalan sebagai berikut dengan besar kesalahan adalah 269.99% yang artinya peramalan tidak akurat.



Gambar 3. Plot data inflasi umum (biru) dan data hasil peramalan (merah)

Selanjutnya, melakukan peramalan untuk 12 data berikutnya atau ramalan inflasi di tahun 2020. Hasil yang diperoleh pada bulan Januari besar inflasi 0.35, bulan Februari -0.08, bulan Maret -0.19, bulan April -0.13, bulan Mei 0.04, bulan Juni 0.34, bulan Juli 0.31, bulan Agustus -0.12, bulan September -0.23, bulan Oktober -0.17, bulan November -0.0, dan bulan Desember 0.29. Hasil peramalan tersebut dapat dibuat plot data sebagai berikut.



Gambar 4. Plot data ramalan bulanan inflasi umum tahun 2020

4. Simpulan

Apabila dilihat pada hasil peramalan dapat diperkirakan besar inflasi terus mengalami naik-turun. Hal ini dikarenakan pada enam periode awal data masa lalu mengalami naik-turun pula. Rata-rata peramalan besar inflasi umum pada tahun 2020 adalah 0,034 dengan besar kesalahan adalah 269,99%.

5. Ucapan Terima Kasih

Tugas ini tidak akan selesai dengan baik tanpa bantuan teman-teman Matematika Pascasarjana IPB angkatan 58 khususnya Fitra Nuvus serta arahan dari M. Najib selaku pembimbing praktikum. Oleh karena itu, penulis ucapkan terima kasih.

6. Daftar Pustaka

- Aswi dan Sukarna. 2006. Analisis Deret Waktu. Makassar : Penerbit Andira.
- Atmadja. (1999). Inflasi di Indonesia: Sumber-sumber Penyebab dan Pengendaliannya. Jurnal Akuntansi dan Keuangan Universitas Kristen Petra, 1(1), 54 – 67.
- Goh, C., & Law, R. 2011. *The methodological progress of tourism demand forecasting: A review of related literature*. Journal of Travel and Tourism Marketing, 28(3), 296-317.
- Hill, Hal. 1996. Transformasi Ekonomi Indonesia Sejak 1966 : Sebuah studi Kritis dan Komprehensif. Yogyakarta: Tiara Wacana Yogyka.
- Makridarkis, et all. 1997. *Forecasting Methods and Applications*. New Jersey: Wiley.

- Makridarkis, et all. 1999. Metode dan Aplikasi Peramalan Edisi Ke-I. Jakarta : Erlangga.
- Mankiw, G. N. 2006. *Principles of Economics*: Pengantar Ekonomi Makro(Ed. 3). Jakarta: Salemba Empat.
- Triwahyuni. 2021. Pengendalian Inflasi, Moneter, dan Fiskal dalam Perspektif Ekonomi Makro Islam. *Ekonomica Sharia: Jurnal Pemikiran dan Pengembangan Ekonomi Syariah*, 6(2), 199-210.

Analisis ARIMA untuk Menentukan Model Peramalan Tingkat Inflasi

Anggun Praptaningsih

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: praptaanggun@apps.ipb.ac.id

Abstrak.

Tingkat inflasi merupakan salah satu faktor penting dalam pengukuran tingkat perekonomian yang berdampak pada kesejahteraan masyarakat. Banyak metode dalam statistika yang dapat digunakan untuk meramalkan tingkat inflasi masa depan dengan harapan dapat dipergunakan oleh pemerintah dalam pengambilan keputusan, salah satunya adalah metode ARIMA. Metode ini merupakan salah satu metode untuk data *time series* yang dapat diterapkan untuk mencari model peramalan terbaik. Pada makalah ini, akan dianalisa dan ditentukan model terbaik dari metode ARIMA yang cocok untuk meramalkan data inflasi pada salah satu IHK yaitu sandang. Hasil Analisa menunjukkan bahwa model ARIMA(0,1,1) merupakan model terbaik untuk melakukan peramalan data inflasi sandang. Hal tersebut disimpulkan dari nilai keakuratan model menggunakan nilai AIC (Akaike Information Criteria). Nilai AIC dari ARIMA(1,1,0) adalah 1106,7 sedangkan nilai AIC dari ARIMA(0,1,1) adalah 184. Dengan ketentuan bahwa semakin kecil nilai AIC, maka semakin baik model yang dipilih. Dengan demikian, model ARIMA(0,1,1) merupakan model terbaik.

Kata Kunci: ARIMA, *time series*, peramalan.

1. Pendahuluan

Dalam teori makro ekonomi, masalah yang sering dihadapi di negara berkembang antara lain adalah permasalahan pertumbuhan ekonomi yang lambat, masalah ketidakstabilan kegiatan ekonomi, kemiskinan, kenaikan harga-harga atau inflasi. Tidak dapat dipungkiri bahwa permasalahan inflasi merupakan masalah yang selalu menjadi perhatian penting bagi negara-negara di dunia khususnya negara berkembang seperti Indonesia. Secara definisi dari Bank Indonesia, inflasi merupakan kecenderungan kenaikan harga pada barang dan jasa yang terus-menerus dan dalam kurun waktu tertentu. Pengukuran tingkat inflasi dilakukan dengan menggunakan data harga dari kelompok pengeluaran pada Indeks Harga Konsumen (IHK). Indeks Harga Konsumen merupakan indeks yang menghitung rata-rata perubahan harga dari barang dan jasa yang dikonsumsi oleh rumah tangga dalam kurun waktu tertentu. Pengelompokan IHK berdasarkan *the Classification of Individual Consumption by Purpose* (COICOP) meliputi 7 kelompok pengeluaran, yaitu 1) Bahan makanan, 2) Makanan jadi, minuman, dan tembakau, 3) Perumahan, 4) Sandang, 5) Kesehatan, 6) Pendidikan dan olahraga, dan 7) Transportasi serta komunikasi.

Berdasarkan data inflasi yang diperoleh dari Badan Pusat Statistik, dalam kurun waktu 15 tahun terakhir, terjadi perubahan tingkat inflasi yang fluktuatif. Hal ini disebabkan dari perekonomian dan tingkat penjualan barang dan jasa suatu negara yang tidak stabil, sehingga menimbulkan inflasi yang dapat memicu terjadinya krisis ekonomi ringan hingga berat. Oleh karena itu, perlu adanya model peramalan yang tepat terhadap tingkat inflasi di Indonesia, agar pemerintah dapat mengambil kebijakan untuk menaikkan pertumbuhan ekonomi dan menghindari krisis ekonomi.

Tujuan ditulisnya makalah ini adalah untuk menentukan model peramalan yang tepat meramalkan tingkat inflasi menggunakan metode ARIMA. Hal ini ditujukan dengan harapan data peramalan dapat dipergunakan oleh pemerintah dalam pengambilan kebijakan untuk kesejahteraan masyarakat dalam bidang ekonomi.

2. Data dan Metode

Data yang digunakan adalah data runtun waktu inflasi yang memuat 7 kelompok IHK dan umum sejak tahun 2005 hingga 2019. Metode yang digunakan adalah metode ARIMA.

2.1. Alat dan Bahan

Alat dan bahan yang digunakan dalam proses pembuatan makalah ini adalah sebagai berikut.

Tabel 1. Spesifikasi Personal Computer (PC)

Aspek	Deskripsi
Operating System	Windows 10 Home Single Language 64-bit (10.0, Build 19043)
System Model	ASUS X441NA
Processor	Intel® Cleron® CPU N3350 @ 1.10GHz (2 CPUs)
Memory	4096 MB RAM
DirectX Version	DirectX 12

Tabel 2. Software JULIA versi 1.6.2 dengan beberapa paket yang digunakan

Paket	Kegunaan
Plots.jl	digunakan untuk membuat grafik
CSV.jl	digunakan untuk membaca file dalam bentuk CSV
DataFrame.jl	digunakan untuk membuat tabulasi data
VMLS.jl	digunakan untuk mengaktifkan fungsi <i>Vectors, Matrices, and Least Square</i>
Statictics.jl	digunakan untuk mengaktifkan fungsi statistika
Turing.jl	mengaktifkan berbagai metode inferensi berbasis pengambilan sampel

2.2. Data

Data inflasi yang diperoleh penulis dalam makalah ini adalah data sekunder yang berasal dari laman Kementerian Perdagangan (<https://statistik.kemendag.go.id/inflation>). Data sekunder merupakan data yang diperoleh dengan cara membaca, mempelajari, dan memahami melalui media lain (Sugiyono, 2012). Data inflasi ini terdiri dari 7 kelompok IHK dan umum. Dalam makalah ini, dipilih salah satu data IHK yaitu data inflasi sandang karena data terlihat sangat fluktuatif dibandingkan dengan data lainnya. Kemudian data inflasi sandang dibagi menjadi 2 yaitu 80% sebagai data training dan 20% sebagai data testing. Data training digunakan untuk melatih algoritma, sedangkan data testing ini akan melihat kinerja algoritma yang sudah dilatih sebelumnya ketika menemukan data baru yang belum pernah dilihat (Olla, 2017).

2.3. Autoregressive Integrated Moving Average Method (ARIMA)

Metode ARIMA juga dikenal sebagai metode runtut waktu Box-Jenkins. Metode ini merupakan gabungan antara model Autoregressive (AR) dan Moving Average (MA) yang dikembangkan oleh George Box dan Gwilyn Jenkins (Wulandari *et al*, 2019). Terdapat beberapa tahapan yang dilakukan dalam metode ARIMA, yaitu menguji stasioner data dan differencing, analisis menggunakan PACF plots untuk menentukan parameter AR dan MA, menentukan model dan sampling, dan menentukan model yang paling baik dengan menghitung AIC. Berikut ini karakteristik metode Box-Jenkins (ARIMA) yang dibagi menjadi 3 kelompok, yaitu AR, MA, dan metode campuran keduanya (Hendranata, 2003).

Autoregressive (AR)

Bentuk model *autoregressive* dengan ordo p ($\text{AR}(p)$) atau model ARIMA $(p, 0, 0)$ dapat dinyatakan sebagai berikut. Misalkan z_1, z_2, \dots adalah sebuah runtut waktu. Model AR untuk runtut waktu memiliki bentuk

$$\hat{z}_{t+1} = \theta_{1z_t} + \dots + \theta_{Mz_{t-M+1}}, \quad t = M, M+1, \dots$$

M adalah *memory of lag* dari model tersebut. \hat{z}_{t+1} adalah prediksi z_{t+1} dibuat pada waktu t (ketika z_t, \dots, z_{t-M+1} diketahui). Prediksi ini merupakan sebuah fungsi linear dari nilai M data series sebelumnya. Dengan pemilihan parameter yang baik, model AR dapat digunakan untuk memprediksi nilai selanjutnya pada sebuah runtut waktu, saat ini maupun nilai M sebelumnya.

Least square (atau regresi) dapat digunakan untuk memilih parameter model AR, berdasarkan data yang telah diobservasi dengan meminimalkan jumlah kuadrat dari *prediction errors*, yakni $z_t - \hat{z}_t$ dengan $t = M+1, \dots, T$.

$$(z_{M+1} - \hat{z}_{M+1})^2 + \dots + (z_T - \hat{z}_T)^2$$

Model AR dapat dibuat menjadi bentuk umum sebagai berikut.

$$y^{(i)} = z_{M+i}, \quad x^{(i)} = z_{M+i-1}, \dots, z_i, \quad i = 1, \dots, T-M.$$

Model ini dapat diidentifikasi melalui nilai ACF yang berpola *dying down* (turun eksponensial atau sinusoidal menuju 0 dengan bertambahnya k) dan nilai bernilai PACF yang *cut off after lag* (terpotong setelah beberapa lag)

Moving Average (MA)

Bentuk umum model moving average ordo q ($\text{MA}(q)$) atau ARIMA $(0, 0, q)$ dinyatakan sebagai berikut.

$$X_t = \mu' + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

Dengan μ' = suatu konstanta, θ_1 sampai θ_q adalah parameter-parameter moving average, e_{t-k} nilai kesalahan pada saat $t-k$. Kebalikan dari model AR, model ini dapat diidentifikasi melalui nilai ACF yang berpola *cut off after lag* dan PACF *dying down*.

ARMA

Model umum untuk campuran proses AR(1) murni dan MA(1) murni, misal ARIMA $(1, 0, 1)$ dinyatakan sebagai berikut:

$$X_t = \mu' + \phi_1 X_{t-1} + e_t - \theta_1 e_{t-1}$$

atau

$$(1 - \phi_1 B)X_t = \mu' + (1 - \theta_1 B)e_t$$

AR(1) MA(1)

Model ini dapat diidentifikasi melalui nilai ACF yang berpola *dying down* dan pola nilai PACF *dying down*.

ARIMA

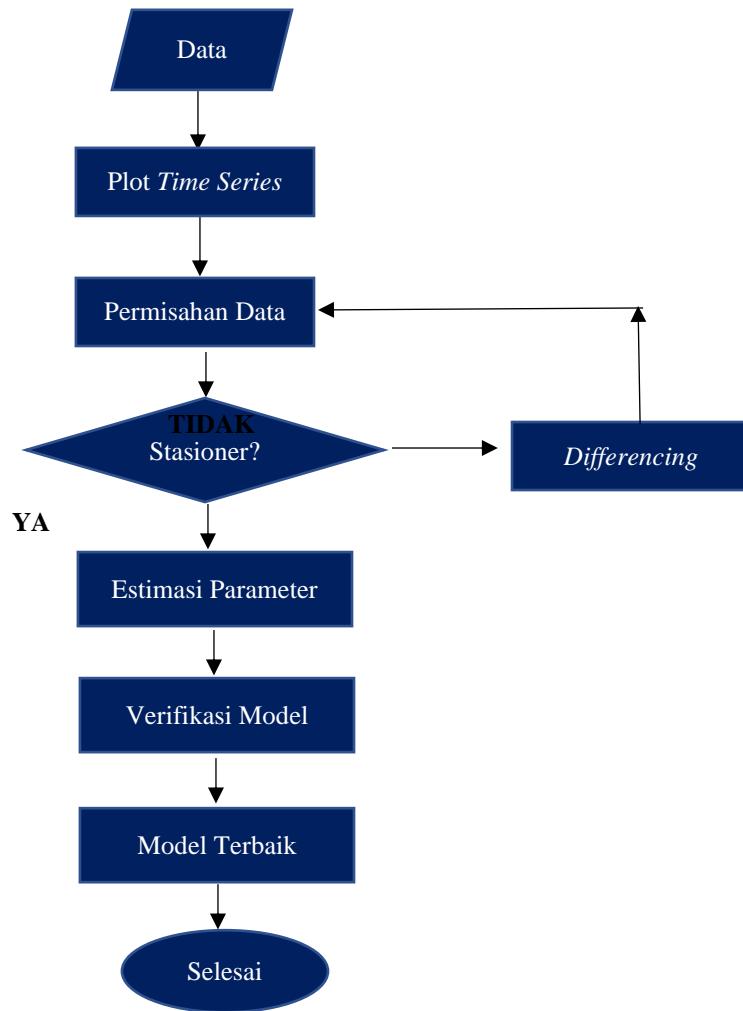
Apabila ketidakstasioneritas ditambahkan pada campuran proses ARMA, maka model umum ARIMA (p, d, q) terpenuhi. Persamaan kasus sederhana ARIMA (1, 1, 1) adalah sebagai berikut.

$$(1 - B)(1 - \phi_1 B)X_t = \mu' + (1 - \theta_1 B)e_t$$

Pembedaan pertama AR(1) MA(1)

Bentuk umum dari ARIMA (p, d, q) dengan p adalah nilai AR, q adalah nilai MA, dan d adalah ordo dari *differencing*.

Langkah-langkah dalam penentuan model terbaik untuk peramalan data ini tergambar dalam *flowchart* berikut.



Gambar 1. Flowchart

Langkah pertama yaitu memasukan data. Dalam makalah ini, data yang akan digunakan adalah data inflasi pada sandang tahun 2005 hingga 2019. Kemudian data tersebut dibuat plot.

Langkah selanjutnya adalah pemisahan data menjadi data *training* dan data *testing*. Data *training* adalah data yang dilatih untuk membuat prediksi atau menjalankan algoritma, sedangkan data *testing* adalah data yang akan dites untuk melihat keakuratan dari algoritmanya (Herlambang, 2018).

Setelah pemisahan data, kemudian dilakukan uji stasioner data menggunakan test *Augmented Dickey-Fuller* (ADF). Hipotesis uji stasioner data menggunakan ADF ini adalah

$$H_0 : \text{data tidak stasioner}$$

$$H_1 : \text{data stasioner}$$

Keputusan diterima atau ditolaknya hipotesis bergantung dari p-value, jika p-value < 0,05 (tingkat signifikansi), maka H_0 ditolak. Jika setelah diuji ternyata data tidak stasioner, maka harus dilakukan *differencing* pada data tersebut dan dilakukan uji stasioner ulang hingga data menjadi stasioner. Banyaknya *differencing* yang dilakukan akan menjadi ordo dari d pada model ARIMA(p, d, q).

Langkah berikutnya adalah estimasi parameter. Penentuan parameter AR dan MA dapat dilakukan dengan plot ACF dan PACF. ACF digunakan untuk membentuk model (p, d, q) sebagai orde dari p dan PACF digunakan untuk membentuk model (p, d, q) sebagai orde dari q . Pada langkah ini, akan ditentukan kandidat model yang akan diverifikasi pada langkah selanjutnya.

Langkah terakhir adalah verifikasi model dengan pengujian residu dan melihat keakuratan model dengan nilai AIC (*Akaike Information Criteria*). Pengujian residual dari data dianggap baik ketika residual bersifat *white noise* dan berdistribusi normal (Isnawaty, 2017). Uji ini menggunakan sampling dengan menggunakan NUTS (No U Turn Sampler). NUTS merupakan teknik sampling yang lebih cepat dibandingkan dengan sampling lainnya. Selanjutnya, model-model perkiraan dilihat keakuratannya menggunakan nilai AIC. Model yang paling baik adalah model yang memiliki nilai AIC paling kecil.

3. Hasil dan Pembahasan

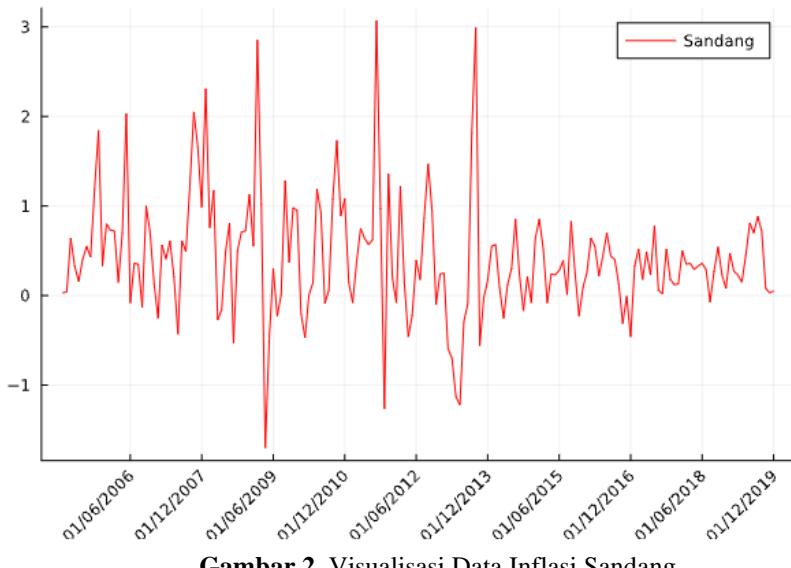
Penentuan model untuk peramalan data inflasi sandang dari tahun 2005 hingga tahun 2019 dengan metode ARIMA dilakukan dalam beberapa tahapan.

Pertama, kita masukan dan visualisasikan data inflasi sandang yang akan kita analisis menggunakan algoritma

Kode 1. Definisi data dan plot data

```
Inflasi = CSV.read("Data_3_Inflasi.csv", DataFrame)
d = Inflasi[:,5]
plot(x,Inflasi[:,5], label="Sandang", xrotation=45, color=:red)
```

sehingga akan muncul visualisasi data inflasi sandang berikut.



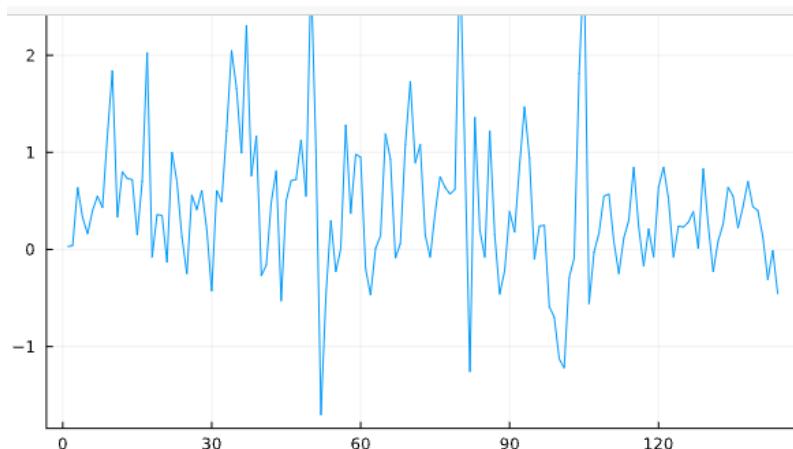
Gambar 2. Visualisasi Data Inflasi Sandang

Dari plot di atas, terlihat kenaikan dan atau penurunan inflasi secara drastis pada akhir tahun 2009, 2012, dan 2013.

Kemudian, kita pisahkan dari menjadi data *training* dan data *testing* dengan mengambil 80% data sebagai data *training*. Berikut ini algoritma pemisahan data dan hasil keluarannya.

Kode 2. Pemisahan data *training* dan data *testing*

```
train_percentage = 0.85
d_train = d[1:floor(Int, train_percentage*length(d))]
N = length(d_train)
# visualisasi data training
plot(d_train)
```

**Gambar 3.** Visualisasi Data Training

Dari plot di atas, terlihat data training dengan persentase 80% dari 180 data (144 data).

Langkah ketiga yaitu mengecek stasioneritas menggunakan Augmented Dickey-Fuller (ADF) test dengan algoritma dan keluaran sebagai berikut.

```
In [104]: ADFTest(d_train, Symbol("constant"), 12)
Out[104]: Augmented Dickey-Fuller unit root test
-----
Population details:
    parameter of interest: coefficient on lagged non-differenced variable
    value under h_0:          0
    point estimate:         -0.723435

Test summary:
    outcome with 95% confidence: fail to reject h_0
    p-value:                  0.1101

Details:
    sample size in regression:      131
    number of lags:                 12
    ADF statistic:                -2.52229
    Critical values at 1%, 5%, and 10%: [-3.4766 -2.88183 -2.57759]
```

Gambar 4. Uji Stasioneritas

Dari data keluaran di atas, terlihat bahwa H_0 diterima, artinya data inflasi sandang tidak stasioner. Oleh karena itu, perlu dilakukan *difference* untuk menghilangkan stasioneritas dan mengecek ulang menggunakan ADF test. Berikut hasil dari *differencing*.

```
In [81]: ADFTest(d_diff, Symbol("constant"), 12)
Out[81]: Augmented Dickey-Fuller unit root test
-----
Population details:
    parameter of interest: coefficient on lagged non-difference
d variable
    value under h_0:          0
    point estimate:         -7.2731

Test summary:
    outcome with 95% confidence: reject h_0
    p-value:                   <1e-06

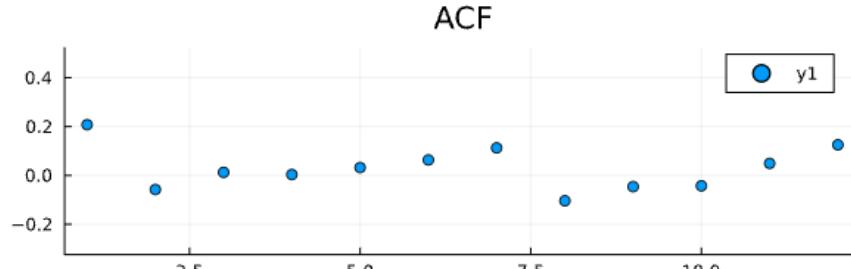
Details:
    sample size in regression:      130
    number of lags:                 12
    ADF statistic:                -5.86764
    Critical values at 1%, 5%, and 10%: [-3.47693 -2.88197 -2.577
```

67]

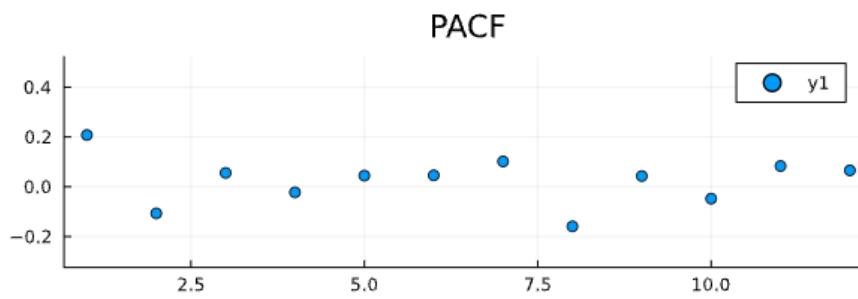
Gambar 5. Data Differencing

Dari penjelasan di atas, kita sudah dapat mengasumsikan bahwa data sudah stasioner dalam mean (karena H₀ ditolak) dengan p-value sangat rendah.

Langkah selanjutnya adalah menentukan parameter AR dan MA dengan bantuan plot ACF (*Auto Correlation Function*) dan PACF (*Partial Auto Correlation Function*). ACF mengindikasikan nilai *Autoregressive* (AR), sedangkan PACF mengindikasikan nilai *Moving Average* (MA). Penentuan data yang mengandung *Autoregressive* (AR) atau *Moving Average* (MA) adalah dengan melihat pola dari ACF dan PACFnnya.



Gambar 6. ACF Data



Gambar 7. PACF Data

Dilihat dari gambar 6 dan gambar 7, terlihat bahwa kandidat model untuk peramalan yaitu:

Berdasarkan plot ACF → ARIMA (0,1,1)

Berdasarkan plot PACF → ARIMA (1,1,0)

Agar penentuan model terbaik memiliki tingkat kepercayaan tinggi, kita uji dua model ARIMA yaitu ARIMA (1, 1, 0) dan ARIMA (0, 1, 1). Model ARIMA (1,1,0) didefinisikan sebagai berikut.

Kode 3. Definisi model ARIMA(1,1,0)

```
@model ARIMA110(x) = begin
    T = length(x)
    μ ~ Uniform(-10, 10)
    φ ~ Uniform(-1, 1)
    for t in 3:T
        val = μ +
            # Drift term.
            x[t-1] +      # ARIMA(0,1,0) portion.
            φ * (x[t-1] - x[t-2]) # ARIMA(1,0,0) portion.
        x[t] ~ Normal(val, 1)
    end
end
```

Dari algoritma di atas, x adalah deret waktu asli dan nilai AR yang dimiliki atau $p = 1$. Berikut adalah algoritma dari ARIMA (0,1,1).

Kode 4. Definisi model ARIMA(0,1,1)

```
@model ARIMA011(x) = begin
```

```
    T = length(x)
```

```
# Mengatur vektor error.
```

```

 $\epsilon = \text{Vector}(\text{undef}, T)$ 
 $x_{\hat{}} = \text{Vector}(\text{undef}, T)$ 

 $\theta \sim \text{Uniform}(-5, 5)$ 

# Treat the first  $x_{\hat{}}$  as a parameter to estimate.
 $x_{\hat{}}[1] \sim \text{Normal}(550, 220)$ 
 $\epsilon[1] = x[1] - x_{\hat{}}[1]$ 

for t in 2:T
    # Memprediksi nilai dari x.
     $x_{\hat{}}[t] = x[t-1] - \theta * \epsilon[t-1]$ 
    # Menghitung observed error.
     $\epsilon[t] = x[t] - x_{\hat{}}[t]$ 
    # Mengobservasi likelihood.
     $x[t] \sim \text{Normal}(x_{\hat{}}[t], 1)$ 
end
end

```

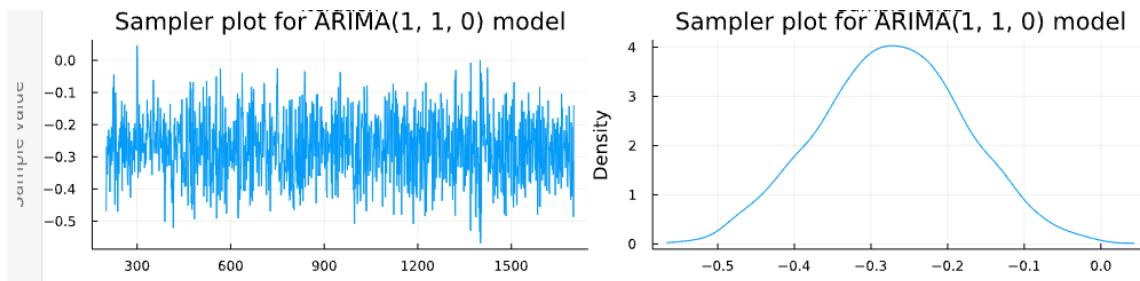
Sama seperti definisi algoritma sebelumnya, x adalah deret waktu asli dan nilai MA yang dimiliki atau $q = 1$. Setelah dua model tersebut didefinisikan algoritmanya, selanjutnya akan dilihat residualnya. Berikut algoritma dari sampling dengan NUTS.

Kode 5. Sampling NUTS

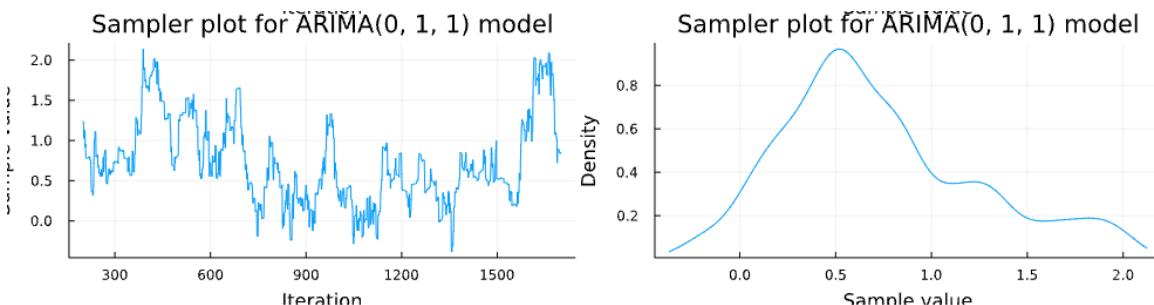
```

chain_ARIMA011 = sample(ARIMA011(d_train), NUTS(200, 0.95), 1500 )
plot(chain_ARIMA011, title = "Sampler plot for ARIMA(0, 1, 1) model")
chain_ARIMA110 = sample(ARIMA110(d_train), NUTS(200, 0.95), 1500 )
plot(chain_ARIMA110, title = "Sampler plot for ARIMA(1, 1, 0) model")

```



Gambar 8. Uji Residual ARIMA (1,1,0)



Gambar 9. Uji Residual ARIMA (0,1,1)

Dari gambar 8 dan 9, terlihat bahwa residual keduanya *white noise* dan berdistribusi normal. Langkah terakhir dalam analisis ini adalah menguji keakuratan dua model tersebut menggunakan perhitungan AIC. Untuk model ARIMA(1,1,0)

diperoleh nilai 1106,7, sedangkan nilai AIC untuk model ARIMA(0,1,1) adalah 184. Diketahui bahwa yang semakin kecil nilai AIC, maka model yang dipilih semakin akurat.

Karena nilai AIC ARIMA (1,1,0) lebih besar daripada ARIMA (0,1,1), maka dapat disimpulkan bahwa model terbaik dan tepat untuk meramal data inflasi sandang adalah ARIMA(0,1,1).

4. Simpulan

Dari hasil analisis data inflasi sandang, diperoleh bahwa model yang cocok dan tepat untuk peramalan data inflasi sandang adalah ARIMA(0,1,1). Hal ini ditunjukkan dari plot ACF dan PACF, serta uji keakuratan menggunakan nilai AIC (*Akaike Information Criteria*). Nilai AIC dari ARIMA (1,1,0) lebih besar dibandingkan dengan nilai AIC dari ARIMA(0,1,1).

5. Ucapan Terima Kasih

Makalah dan mini projek ini dapat terselesaikan salah satunya oleh bantuan dari beberapa pihak. Penulis mengucapkan terima kasih kepada Fitra Nuvus, Refi, Rahmawati, Trianty, Ade, dan seluruh pihak dan tidak bisa disebutkan satu per satu.

6. Daftar Pustaka

- Bank Indonesia.[diakses 1 Oktober 2021]. <https://www.bi.go.id/fungsi-utama/moneter/inflasi/default.aspx>
- Hendranata A. 2003. ARIMA (Autoregressive Moving Average), Manajemen Keuangan Sektor Publik FEUI.
- Herlambang MB. 2018. Training dan Test Set. [diakses 18 Oktober 2021]. <https://www.megabagus.id/training-set-test-set/>
- Isnarwaty DP. 2017. Analisis Time Series Menggunakan ARIMA Box-Jenkins. [diakses 18 Oktober 2021]. https://rstudio-pubs-static.s3.amazonaws.com/453202_bca92246b21a40a9bfc98f83ec880a83.html
- Olla K. 2017. Belajar Machine Learning dalam Pengolahan Data. [diakses 17 Oktober 2021]. <https://www.jagoanhosting.com/blog/belajar-machine-learning-dalam-pengolahan-data-ini-panduannya/>
- Rufaidah A, Effindi MA. 2018. Analisis Time Series untuk Menentukan Model Terbaik Produk Songkok Nasional di Kabupaten Gresik. Di dalam: Prosiding Seminar Nasional Matematika dan Terapannya 2018. p-ISSN : 2550-0384.
- Rumus Statistik. ___. Uji Stasioneritas ADF dengan R. [diakses 18 Oktober 2021]. <https://www.rumusstatistik.com/2021/05/uji-stasioner-augmented-dickey-fuller-r.html>
- Saumyaghshah. 2019. Improved ARIMA Implementation. [diakses 16 Oktober 2021]. https://github.com/saumyaghshah/JupyterNBTuringExamples/blob/master/time_series_ARIMA.ipynb
- Sugiyono. 2012. Memahami Penelitian Kualitatif. Bandung: ALFABETA
- Sumber Data Inflasi. [diakses 1 Oktober 2021]. <https://www.bps.go.id/subject/3/inflasi.html>
- Wulandari R, Rahmat G. 2019. Metode *Autoregressive Integrated Movingaverage Arima* dan Metode *Adaptive Neuro Fuzzy Inference System* (Anfis) dalam Analisis Curah Hujan. Berkala Fisika, Vol. 22, No. 1, Januari 2019, Hal 41-48.

Analisis Faktor yang Mempengaruhi Cuaca di Jakarta Menggunakan Metode *Principal Component Analysis (PCA)*

Yusti Qomah

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: gomahyq25yusti@apps.ipb.ac.id

Abstrak.

Cuaca adalah hal yang tidak bisa diprediksi secara manual, selalu berubah setiap waktu. Perubahan cuaca dipengaruhi oleh berbagai macam faktor, seperti suhu, kecepatan angin, kelembapan, tekanan udara, dan curah hujan. Untuk menganalisis faktor yang mempengaruhi cuaca dengan data berdimensi besar dapat menggunakan metode *Principal Component Analysis (PCA)*. Makalah ini menggunakan metode PCA untuk menganalisis faktor yang mempengaruhi cuaca di Jakarta pada tahun 2019. Variabel dalam data asli yang berupa *timestamp, temperature, precipitation total, relative humidity, wind speed, dan wind direction* akan direduksi menjadi 2 dimensi baru tanpa menghilangkan informasi dari data asli. *Principal component (PC)* yang dihasilkan dapat menjelaskan total 79% informasi dari data asli, dengan PC 1 menjelaskan informasi keseluruhan data sebanyak 59% dan PC 2 dapat menjelaskan informasi keseluruhan data sebanyak 20%. Error yang dihasilkan dari matriks yang dibentuk ulang berdasarkan PC dan matriks asli adalah 19,618524437004883. Varians kumulatif yang dihasilkan masing-masing PC berada antara 0,5 dan 1. Pola komponen utama yang dihasilkan menyerupai pola data temperature dan wind speed. Dapat disimpulkan, bahwa kedua PC yang dihasilkan dari metode PCA dapat menjelaskan informasi secara keseluruhan dengan baik.

Kata Kunci: Principal Component Analysis, cuaca, reduksi dimensi.

1. Pendahuluan

Sebelum teknologi semakin berkembang, metode statistik yang sering digunakan juga masih sederhana. Penerapan analisis multivariat belum banyak dilakukan. Seiring berkembangnya teknologi, analisis multivariat diaplikasikan pada berbagai macam masalah dari berbagai disiplin ilmu. Metode dalam analisis multivariat seperti *Linear Discriminant Analysis (LDA)* dan *Principal Component Analysis (PCA)* membutuhkan banyak perhitungan matematis yang memerlukan bantuan komputasi, tidak bisa diselesaikan secara manual (Santoso, 2018).

Perkembangan teknologi berbanding lurus dengan berbagai macam fenomena alam yang terjadi. Penyebab terjadinya fenomena tersebut salah satunya adalah cuaca. Cuaca dapat diartikan sebagai suatu keadaan atau suasana yang diartikan dalam ruang udara (Endarmoko, 2007). Banyak hal yang mempengaruhi kondisi cuaca di suatu tempat, seperti suhu, kecepatan angin, kelembapan, tekanan udara, dan curah hujan (Yuniar dkk, 2013). Jakarta sebagai ibukota Indonesia menjadi daerah mengalami banyak fenomena alam, khususnya bencana. Kondisi cuaca yang buruk dapat berakibat buruk jika tidak diantisipasi lebih awal. Hal ini disebabkan oleh perbedaan kondisi cuaca (Krishnamurthi dkk, 2015).

PCA adalah metode yang digunakan untuk mentransformasikan variabel yang saling berhubungan dalam sebuah data menjadi data baru yang variabelnya tidak saling berhubungan (Nong, 2008). Jika sebuah data dianalisis secara langsung, jumlah data yang banyak dapat menghambat proses pengolahan data. Untuk mengatasi hambatan tersebut, perlu dikakukan reduksi terhadap variabel yang terdapat di dalam data tanpa menghilangkan informasi dari keseluruhan data. Analisis komponen utama adalah metode yang tepat untuk melakukan analisis multivariat dan visualisasi pada data yang berdimensi besar (Datta *et al.* 2017).

Proses transformasi dalam metode PCA tetap mempertahankan keberagaman yang dapat dijelaskan (Johnson and Wicern, 2002). Cheng He (2016) menunjukkan bahwa analisis data interval-nilai menggunakan PCA untuk mengetahui variabel dengan data interval di bawah musim dapat menampilkan lebih banyak informasi. Peramalan cuaca juga dapat diprediksi menggunakan PCA yang menghasilkan kesalahan prediksi rata-rata dan rata-rata persentase kesalahan prediksi adalah 1,41°C dan 7,93% (Jaruszewicz, 2002). Penelitian pada data curah hujan musiman di U.S. dan China juga pernah dilakukan dengan membandingkan metode PCA dan *Directional Component Anaysis (DCA)*. Penelitian ini menunjukkan bahwa hasil analisis menggunakan PCA lebih baik dalam menjelaskan informasi yang terkandung dalam data (Jewson, 2020). Penelitian lain yang menerapkan PCA dalam penurunan skala statistik-dinamis menunjukkan adanya pengurangan yang hampir mencapai 30% dari kesalahan rata-rata prediksi setelah sebelumnya menggunakan metode tradisional (Arroyo *et al.* 2013).

Makalah ini menganalisis faktor apa yang mempengaruhi cuaca di Jakarta pada tahun 2019 dengan melakukan reduksi pada variabel-variabel dalam data cuaca menggunakan metode PCA. Dari seluruh variabel, akan dilihat variabel

apa saja yang mempengaruhi, akan diketahui variabel yang memiliki pengaruh besar, dan akan dilakukan visualisasi terhadap data cuaca di Jakarta pada tahun 2019.

2. Data dan Metode

2.1. Alat dan Bahan

Hardware dan software yang digunakan dalam makalah ini dapat dilihat pada Tabel 1:

Tabel 1. Alat dan bahan yang digunakan

No.	Nama Alat dan Bahan	Spesifikasi
1.	Prosesor Intel (R) Core(TM)	i3-6006U @ 2,00GHz
2.	Memori	4,00GB RAM
3.	Sistem Operasi	Windows 10 64bit
4.	Julia	1.6.2

Paket yang digunakan dalam makalah ini untuk mengolah data dalam Julia 1.6.2 adalah sebagai berikut:

1. XLSX

Paket XLSX digunakan untuk membaca data yang tersimpan dalam file excel yang memiliki ekstensi .xlsx.

2. Plots

Salah satu tolak ukur baik atau tidaknya data yang ditampilkan adalah visualisasi. Visualisasi yang ditampilkan harus menarik dan dapat menunjukkan arti dari data tersebut. Paket Plots digunakan untuk merepresentasikan data dalam bentuk gambar. Paket Plots memiliki beragam jenis plot, contohnya histogram, line plot, dan scatter. Plot yang digunakan dalam makalah ini adalah line plot.

3. DataFrames

Data mentah dalam berbagai jenis ekstensi file harus direpresentasikan kembali di Julia dalam bentuk data frame menggunakan paket DataFrames. Data yang sudah didefinisikan dalam bentuk data frame ini kemudian akan diolah dengan berbagai macam metode.

4. Dates

Paket Dates digunakan untuk mendefinisikan bentuk waktu dari sebuah data waktu. Umumnya, format waktu dalam paket Dates didefinisikan dalam bentuk yyyy-mm-dd.

5. VegaLite

Paket VegaLite menyediakan berbagai macam jenis plot, contohnya *circle*, *area*, dan *point*. Visualisasi yang didefinisikan oleh paket VegaLite adalah format *JavaScript Object Notation (JSON)*.

6. Statistics

Paket Statistics adalah paket dalam Julia yang menyediakan fungsi-fungsi statistik.

7. StatsBase

StatsBase adalah paket yang menyediakan fungsi-fungsi dasar statistik, seperti kovarian, statistik skalar, dan pengambilan sampel.

8. MultivariateStats

MultivariateStats adalah paket yang menyediakan fungsi statistik multivariat seperti *Principal Components Analysis (PCA)*, *Canonical Correlation Analysis (CCA)*, dan *Linear Discriminant Analysis (LDA)*.

9. LinearAlgebra

Paket LinearAlgebra menyediakan fungsi-fungsi dasar dari aljabar linear yang berhubungan dengan matriks, contohnya untuk menghitung norm dan membuat berbagai macam jenis matriks.

10. NaNStatistics

Paket NaNStatistics adalah paket dalam Julia yang menyediakan fungsi movmean. Movmean digunakan untuk melakukan proses *smoothing* pada data menggunakan rata-rata.

2.2. Data

Data yang digunakan pada makalah ini adalah data cuaca di Jakarta per jam yang bersumber dari <https://www.meteoblue.com/>. Rentang waktu dalam data yang digunakan adalah dari Januari 2019 sampai Desember 2019. Variabel-variabel yang terdapat dalam data cuaca tersebut adalah waktu (*timestamp*), *temperature*, *precipitation total*, *relative humidity*, *wind speed*, dan *wind direction*.

2.3. Principal Component Analysis (PCA)

PCA adalah salah satu metode analisis multivariat yang mengolah suatu data dengan cara mereduksi dimensi data agar menghasilkan dimensi data yang lebih sederhana (Jolliffe and Penny, 2001). Dimensi data sederhana yang dihasilkan dari metode PCA disebut *principal component* (PC) (Saepurohman dan Putro, 2019). PC dapat dibentuk dari

matriks kovarian ataupun matriks korelasi. PC dibentuk dari matriks korelasi jika variabel-variabel dari sebuah data memiliki satuan pengukuran yang berbeda, sehingga variabel tersebut harus distandardisasi terlebih dahulu menggunakan persamaan berikut:

$$\text{standardisasi data} = \frac{\text{data tiap kolom} - \text{mean tiap kolom}}{\text{standar deviasi tiap kolom}} \quad (1)$$

Sembarang vektor $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ memiliki matriks kovarian dengan nilai eigen berturut-turut yaitu $\lambda_1 \geq \lambda_1 \geq \dots \geq \lambda_n \geq 0$, maka kombinasi linear yang memenuhi adalah:

$$\begin{aligned} Z_1 &= \mathbf{e}_1^T \mathbf{x} = e_{11}x_1 + e_{21}x_2 + \dots + e_{n1}x_n \\ Z_2 &= \mathbf{e}_2^T \mathbf{x} = e_{12}x_1 + e_{22}x_2 + \dots + e_{n2}x_n \\ &\vdots \\ Z_n &= \mathbf{e}_n^T \mathbf{x} = e_{1n}x_1 + e_{2n}x_2 + \dots + e_{nn}x_n \end{aligned} \quad (2)$$

dengan:

Z_1 = PC ke-1 dengan varians terbesar

Z_2 = PC ke-2 dengan varians terbesar kedua

Z_n = PC ke- n dengan varians terbesar ke- n

x_1 = variabel asal ke-1

x_n = variabel asal ke- n

\mathbf{e}_n = vektor eigen dari data ke- n

(Juaeni, 2014).

Secara umum, model PC ke- i dapat dinyatakan sebagai berikut:

$$Z_i = \mathbf{e}_i^T \mathbf{x}, \quad (3)$$

dimana $i = 1, 2, \dots, n$, sehingga:

$$\text{Var}(Z_i) = \mathbf{e}_i^T \Sigma \mathbf{e}_i = \lambda_i \quad (4)$$

dimana $i = 1, 2, \dots, n$, dan

$$\text{Cov}(Z_i, Z_k) = \mathbf{e}_i^T \Sigma \mathbf{e}_k = 0, i \neq k \quad (5)$$

Jika PCA linier, maka pembentukan PC dapat menggunakan matriks korelasi, dengan:

$$\text{proporsi variansi ke-}i = \frac{\lambda_i}{n} \quad (6)$$

(Johnson and Wichern, 2002).

2.4 Tahapan

1. Mengelompokkan data per jam menjadi data per hari
2. Menginputkan data cuaca per hari yang berformat format .xlsx
3. Mendefinisikan data yang sudah diinput menjadi data frame
4. Mendeskripsikan ukuran-ukuran dari tiap variabel dalam data
5. Merepresentasikan data ke dalam line plot dan violin plot
6. Melakukan analisis dan visualisasi menggunakan metode PCA
7. Menarik kesimpulan

3. Hasil dan Pembahasan

3.1 Deskripsi dan Visualisasi Data

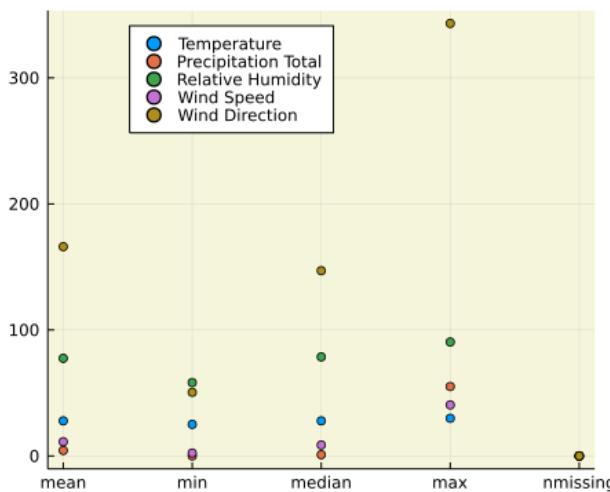
Data yang digunakan adalah data cuaca per jam di Jakarta pada tahun 2019 berformat .xlsx. Setelah data per jam diubah menjadi data cuaca per hari, seluruh data dimasukkan ke dalam Julia dan didefinisikan sebagai *data frame*. Total baris 365 menyatakan jumlah hari pada tahun 2019 dan 6 kolom menyatakan 6 variabel yang ada dalam data tersebut.

Setiap variabel memiliki ukuran pemasaran data. Ukuran pemasaran data dapat dihitung menggunakan fungsi *describe*. Hasil dari ukuran data dapat dilihat dalam Tabel 2 dan dapat divisualisasikan seperti pada Gambar 1.

Tabel 2. Ukuran pemasaran data dari tiap variabel

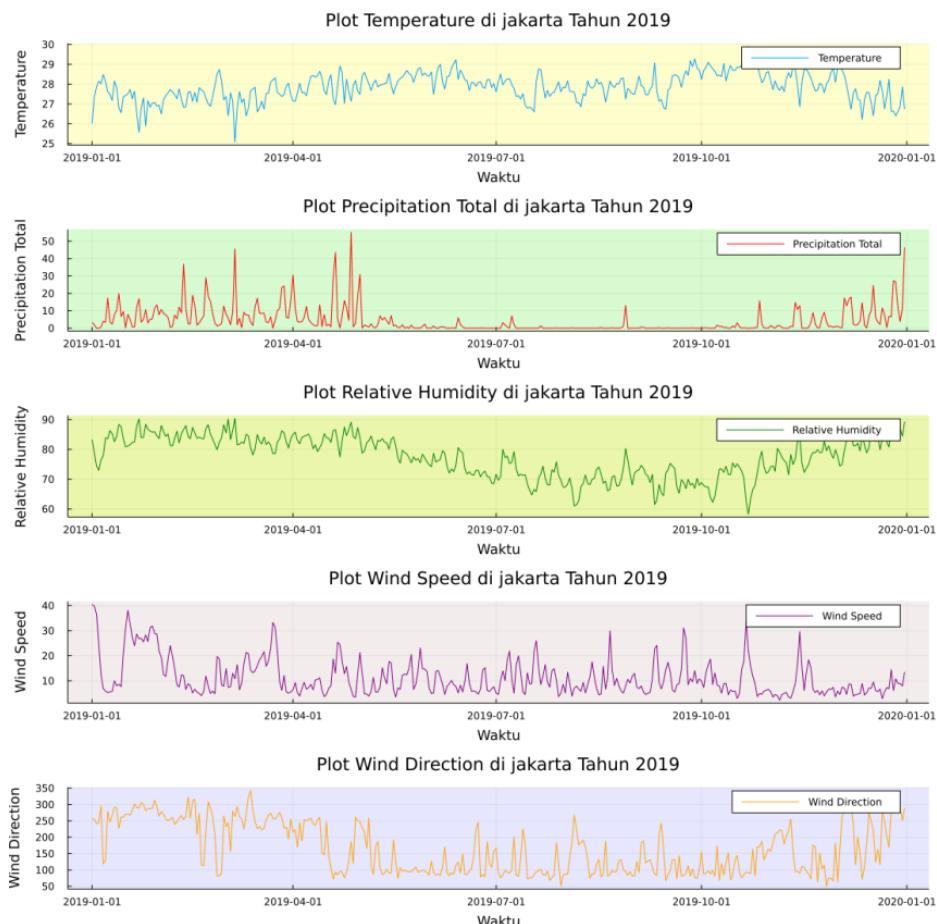
Variabel	Rata-rata	Median	Nilai Minimum	Nilai Maksimum	Data Hilang
Temperature	27,8555	27,8596	25,0729	29,905	0
Precipitation Total	4,36712	1,0	0,0	55,1	0
Relative Humidity	77,4805	78,5991	58,2385	90,4227	0
Wind Speed	11,2828	8,67651	2,28491	40,4807	0
Wind Direction	166,05	146,99	50,4684	343,201	0

Visualisasi Ukuran Pemusatan Data



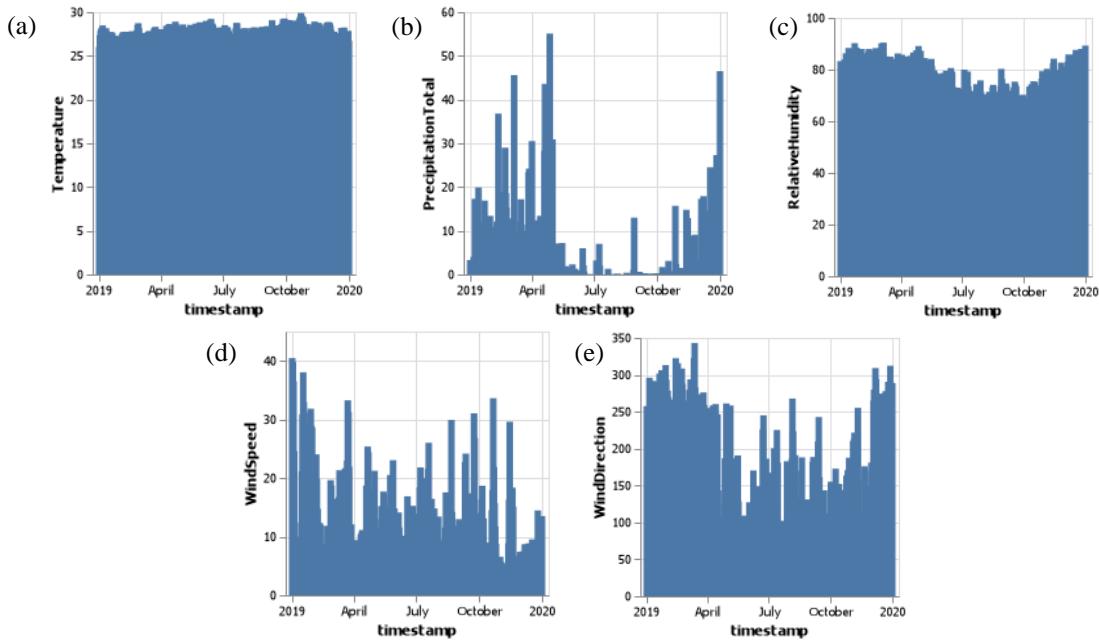
Gambar 1. Scatter plot ukuran pemusatan data dari setiap variabel

Data asli juga dapat divisualisasikan dengan line plot yang menunjukkan pergerakan angka dari tiap variabel per hari dalam tahun 2019. Visualisasi keseluruhan data dapat dilihat pada Gambar 2.



Gambar 2. Line plot seluruh variabel dari data cuaca per hari

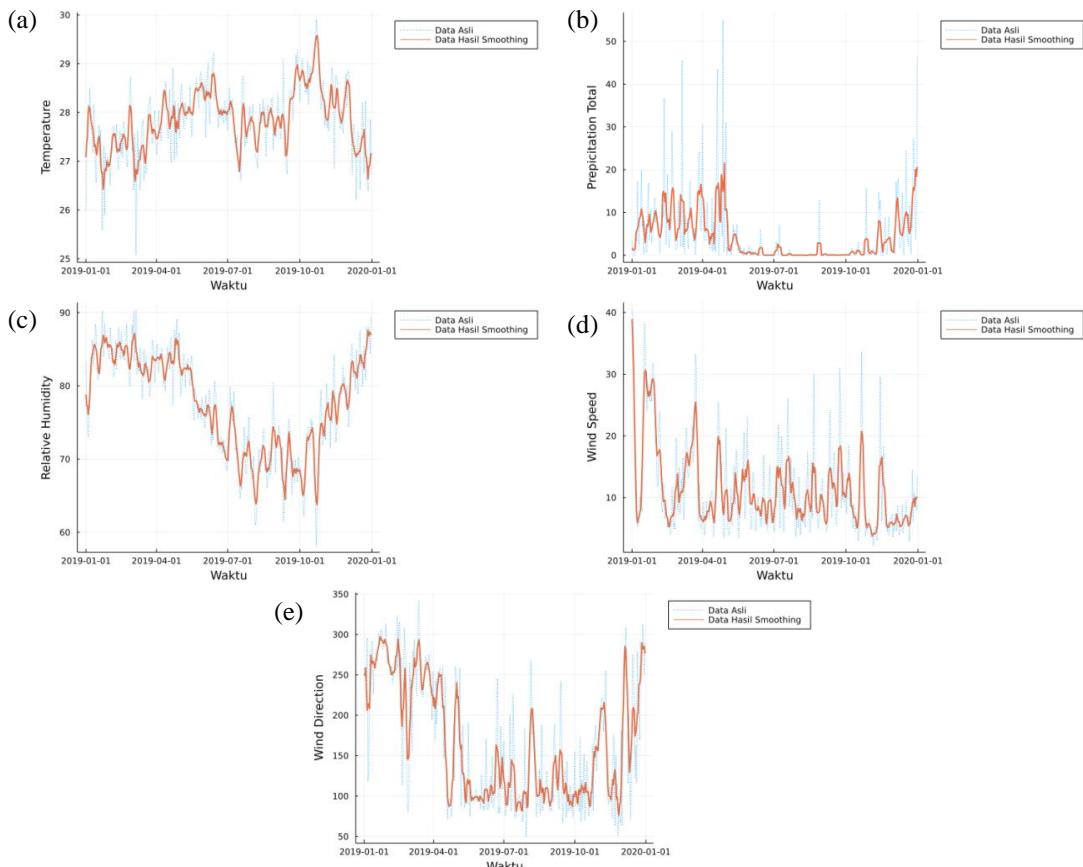
Selain menggunakan line plot, visualisasi data bisa dilakukan menggunakan paket VegaLite. Paket VegaLite menyediakan fungsi untuk membuat plot dalam bentuk *circle*, *area*, *bar*, dan bentuk plot lainnya. Visualisasi data dalam bentuk *bar* dapat dilihat dalam Gambar 3.



Gambar 3. Visualisasi data menggunakan vplot terdiri atas: (a) Plot *Temperature* (b) Plot *Precipitation Total* (c) Plot *Relative Humidity* (d) Plot *Wind Speed* dan (e) Plot *Wind Direction*

3.2 Smoothing Data

Berdasarkan visualisasi data pada Gambar 2 dan Gambar 3 dapat dilihat bahwa *noise* yang dimiliki data tersebut masih relatif tinggi. Agar hasil analisis data yang dihasilkan lebih optimal, perlu dilakukan pengurangan terhadap *noise* data. Salah satu cara melalukan proses smoothing pada suatu data adalah menggunakan rata-rata. Dalam Julia, data cuaca yang sudah dimasukkan diperhalus menggunakan fungsi `movmean` dalam paket `NanStatistics`. Hasil proses *smoothing* ditunjukkan oleh visualisasi berikut:



Gambar 4. Visualisasi hasil proses *smoothing* data terdiri atas: (a) Plot *Temperature* (b) Plot *Precipitation Total* (c) Plot *Relative Humidity* (d) Plot *Wind Speed* dan (e) Plot *Wind Direction*

3.3 Analisis dan Visualisasi Data Menggunakan PCA

Data yang menjadi masukan dalam proses analisis menggunakan PCA adalah data hasil dari proses *smoothing*. Variabel yang tidak mempengaruhi variabel lain perlu dimasukkan ke proses analisis. Pada data cuaca di Jakarta tahun 2019, variabel yang akan dihilangkan adalah *timestamp*. Setelah ada variabel yang dihilangkan, data didefinisikan sebagai *dataN* yang merupakan matriks berukuran 365x5. Matriks *dataN* dinyatakan sebagai berikut:

$$dataN_{365 \times 5} = \begin{bmatrix} 27,0839 & 1,733 & 78,8155 & 38,935 & 251,416 \\ 27,3518 & 1,3 & 77,3618 & 35,3498 & 249,52 \\ 27,4843 & 1,16 & 77,1917 & 30,7655 & 258,834 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 27,1643 & 20,6667 & 86,9412 & 10,1418 & 276,401 \end{bmatrix}$$

Data hasil proses pengurangan variabel menjadi masukan dalam proses standardisasi data. Proses standardisasi dilakukan karena data memiliki satuan dan rentang yang berbeda. Secara manual, standardisasi data dihitung menggunakan Persamaan (1). Penulisan kode standardisasi dalam julia yang menggunakan fungsi *fit* dapat dilihat dalam Kode 1. Hasil proses scaling disimpan kembali dalam *dataN*.

Kode 1. Proses standardisasi data

```
dataN = (dataN .- mean(dataN,dims = 1)) ./ std(dataN,dims=1)
```

$$dataN_{365 \times 5} = \begin{bmatrix} -1,34482 & -0,520443 & 0,208461 & 4,70676 & 1,24525 \\ -0,87867 & -0,606947 & -0,017405 & 4,09666 & 1,2176 \\ 0,647992 & -0,634894 & -0,043827 & 3,31654 & 1,35345 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -1,20489 & 3,2591 & 1,47094 & -0,19301 & 1,60968 \end{bmatrix}$$

Variabel *dataN* dari proses standardisasi data menghasilkan entri-entri baru tanpa merubah ukuran matriks. Matriks *dataN* yang berukuran 365x5 akan diidentifikasi oleh Julia sebagai matriks berukuran 5x365, sehingga perlu dilakukan transpose pada matriks *dataN* yang dinotasikan dengan *dataN'*. Matriks transpose akan menjadi masukan dalam proses reduksi dimensi menggunakan PCA. Data awal yang memiliki 5 variabel akan direduksi menjadi 2 variabel baru. Kedua variabel baru tersebut tidak akan menghilangkan seluruh informasi yang terdapat dalam data awal.

Kode 2. Proses reduksi dimensi menggunakan PCA

```
p = fit(PCA,dataN',maxoutdim=2)
M = projection(p)
principalvars(p)
principalvars(p) ./ tvar(p) * 100
```

Berdasarkan Kode 2, p adalah model PCA dengan dimensi input 5, dimensi output 2, dan principal ratio 0.788523. Nilai principalratio menunjukkan bahwa perintah akan berhenti saat nilai varians kumulatif sudah melewati nilai principalratio tersebut. Variabel M menghasilkan matriks proyeksi berukuran 5x2. Setiap kolom matriks M menyatakan PC. Kolom pertama menyatakan PC1 dan kolom kedua menyatakan PC2. Secara manual, perhitungan model PC dilakukan menggunakan Persamaan (3). Matriks M dinyatakan sebagai berikut:

$$M_{5 \times 2} = \begin{bmatrix} 0,454307 & -0,132883 \\ -0,498009 & -0,275585 \\ -0,508987 & -0,218122 \\ -0,160676 & 0,923488 \\ -0,510595 & 0,0773866 \end{bmatrix}$$

Kontribusi setiap PC dalam menjelaskan informasi yang terkandung di dalamnya dapat dilihat dengan fungsi *principalvars*. PC1 dapat menjelaskan keseluruhan informasi dari data awal sebanyak 2.930657535423525 dan PC 2 dapat menjelaskan informasi sebanyak 1,0119619670892313. Nilai kontribusi dari setiap PC dapat dinyatakan dalam persentase, sehingga PC1 menjelaskan informasi sebanyak 59% dan PC2 dapat menjelaskan informasi sebanyak 20%. Total informasi yang dapat dijelaskan dari PC1 dan PC2 adalah 79%.

Data dalam matriks transpose *dataN'* ditransformasikan menjadi PC atau komponen utama. Hasil transformasi yang dinotasikan sebagai *pcDat* kemudian menjadi masukan dalam proses pembentukan ulang matriks yang mendekati matriks *dataN'* dan dinotasikan sebagai *Xr*. Matriks *pcDat* dan *Xr* dinyatakan sebagai berikut:

$$pcDat_{2 \times 365} = \begin{bmatrix} -1,84996 & -1,36799 & \dots & -3,93474 \\ 4,71966 & 4,16527 & \dots & -1,03974 \end{bmatrix}$$

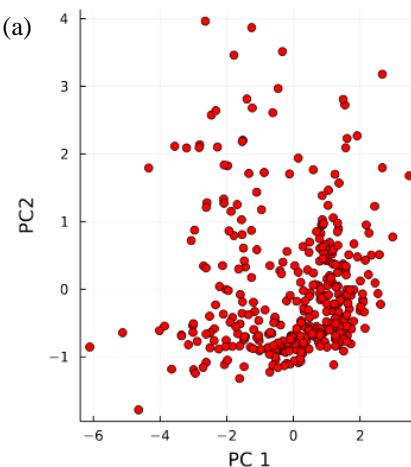
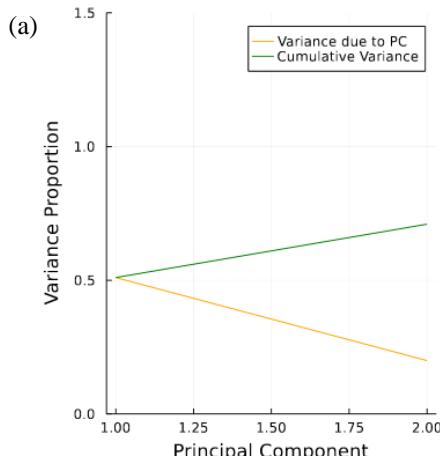
$$Xr_{5 \times 365} = \begin{bmatrix} -1,46761 & -1,17498 & -0,992889 & \dots & -1,53765 \\ -0,379372 & -0,46661 & -0,0359929 & \dots & 2,15423 \\ -0,087854 & -0,21224 & -0,149408 & \dots & 2,13103 \\ 4,65579 & 4,06637 & 3,36467 & \dots & -0,43134 \\ 1,30982 & 1,02083 & 0,868493 & \dots & 1,80822 \end{bmatrix}$$

Pendekatan hasil matriks Xr terhadap matriks $dataN'$ menghasilkan error yang dapat dihitung menggunakan fungsi norm . Error yang dihasilkan adalah 19,618524437004883. Nilai error tidak akan pernah mencapai 0.

Hasil analisis metode PCA perlu divisualisasikan dalam bentuk gambar agar mudah disampaikan dan dipahami. Kode visualisasi hasil metode PCA dalam Julia dapat dilihat pada Kode 4 dan hasil visualisasi dapat dilihat dalam Gambar 5.

Kode 4. Visualisasi hasil PCA

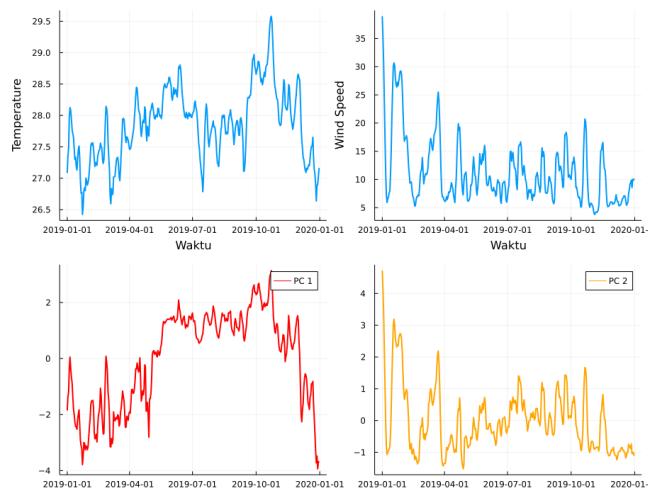
```
pcVar=principalvars(p) ./ tvar(p)
cumVar=cumsum(pcVar)
pcDat=M[:,1:2] '*dataN'
p1=plot(pcVar, c=:orange, label="Variance due to PC")
plot!(1:length(cumVar), cumVar, label="Cumulative Variance", c=:green, xlabel="Principal Component",
      ylabel="Variance Proportion", ylims=(0,1.5))
p2=scatter(pcDat[1,:],pcDat[2,:], c=:red, xlabel="PC 1", ylabel="PC2", msv=0, legend=:none)
plot(p1,p2,size=(700,400))
```



Gambar 5. Visualisasi hasil PCA terdiri atas: (a)line plot PC dan varians kumulatif
(b)scatter plot dari kedua PC

Perhitungan pcVar dilakukan menggunakan Persamaan (6). Gambar 4 (a) menunjukkan varians dari setiap PC. PC 1 memiliki varians 0,5 dan PC2 memiliki varians yang mendekati 0,2. Varians kumulatif yang dihasilkan masing-masing PC berada antara 0,5 dan 1. Gambar 4 (b) menunjukkan hubungan antara PC1 dan PC2 dari matriks hasil transformasi $dataN'$ dengan *principal component*. Setiap titip pada Gambar 4 (b) menyatakan data cuaca di Jakarta setiap hari dalam tahun 2019.

Setelah mendapatkan hasil dari analisis menggunakan metode PCA, plot setiap komponen utama dibandingkan dengan plot setiap variabel dari data yang menjadi masukan saat memulai proses analisis PCA. Perbandingan plot tersebut menunjukkan bahwa pola kedua komponen utama menyerupai pola dari data *temperature* dan *wind speed*. Kemiripan pola tersebut dapat dilihat pada Gambar 6.



Gambar 6. Perbandingan visualisasi komponen utama dengan data *temperature* dan *wind speed*

4. Simpulan

Berdasarkan hasil dan pembahasan, PC 1 dan PC 2 yang dibentuk dari reduksi dimensi menggunakan metode PCA dapat menjelaskan informasi yang terkandung dalam data asli sebanyak 79%. Setiap PC masing-masing dapat menjelaskan informasi sebanyak 59% dan 20% dengan error yang dihasilkan oleh rekonstruksi berdasarkan PC dan matriks asal adalah 19,618524437004883. PC1 memiliki varians 0,5 dan PC 2 memiliki varians 0,2. Pola komponen utama yang dihasilkan menyerupai pola data *temperature* dan *wind speed*. Kedua PC yang dihasilkan dapat menjelaskan dan memvisualisasikan hasil PCA dengan baik tanpa menghilangkan informasi secara keseluruhan.

5. Ucapan Terima Kasih

Terima kasih kepada Kak MK Najib sebagai pengampu praktikum Komputasi Matematika yang telah memberikan bimbingan dan koreksi selama penulisan makalah ini.

6. Daftar Pustaka

- Arroyo RC, Galiana SL, Probst O, Rodrigo JS. 2013. On the application of Principal Component Analysis for Accurate Statistical-Dynamical Downscaling of Wind Fields. *Energy Procedia*. 40: 67-76.
- Datta S. et al. 2017. Multivariate Statistical Data Analysis-Principal Component Analysis (PCA). *International Journal of Livestock Research*. 7(5): 60-78.
- Endarmoko, Eko. 2007. *Tesaurus Bahasa Indonesia*. Jakarta: Gramedia.
- Johnson RA and Wichern DW. 2002. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Santoso S. 2018. *Mahir Statistik Multivariat dengan SPSS*. Jakarta: PT Elek Media Komputindo.
- He, Chong-Cheng and Jen, Jin-Tsong. 2016. Feature Selection of Weather Data with Interval Principal Component Analysis. Di dalam: *International Conference on System Science & Engineering 2016*.
- Jaruszewicz M and Mandziuk J. 2002. Application of PCA method to weather prediction task. Di dalam: *9th International Conference on Neural Information Processing*.
- Jewson, Stephen. 2020. An Alternative to PCA for Estimating Dominant Patterns of Climate Variability and Extremes, with Application to US and China Seasonal Rainfall. *Atmosphere Journal of Multidisciplinary Digital Publishing Institute*. 11: 354 doi:10.3390/atmos11040354.
- Juaeni, Ina. 2014. Dampak Penerapan Principal Component Analysis (PCA) dalam Clustering Curah Hujan di Pulau Jawa, Bali, dan Lombok. *Jurnal Sains Dirgantara*. 11 (2): 97-108.
- Jolliffe, Ian and Penny, KI. 2001. A comparison of multivariate outlier detection methods for clinical laboratory safety data. *Journal of the Royal Statistical Society*. 50 (3): 295-307.
- Krishnamurthi K, Suraj T, Lokesh K, Arum P, 2015. Arduino Based Weather Monitoring System. *International Journal of Engineering Research and General Science*. 3 (2): 452-458.
- Nong, Ji-Fu. The Rainfall Forecast Model of PCA-RBF Neural Networks Based On MATLAB. *Journal of Tropical Meteorology*. 24 (6): 713-717.
- Saepurohman T, Putro BE. 2019. Analisis Principal Component Analysis (PCA) Untuk Mereduksi Faktor-Faktor yang Mempengaruhi Kualitas Kulit Kikil Sapi. Di dalam: *Seminar dan Konferensi Nasional IDEC*.
- Yuniar, RJ., S, DR., Setyawati, O. 2013. Perbaikan Metode Prakiraan Cuaca Bandara Abdulrahman Saleh Dengan Algoritma Neural Network Backpropagation. *Jurnal EECCIS*. 7(1): 65-70.

Klasterisasi Curah Hujan Menggunakan Metode *K-Means* dan *K-Medoids* (Studi Kasus di Jakarta Tahun 2019)

Euis Aprianti

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: aprianti.euis@apps.ipb.ac.id

Abstrak.

Indonesia memiliki dua musim yaitu musim kemarau dan musim penghujan yang keduanya dipengaruhi oleh curah hujan. Sehingga curah hujan dapat memengaruhi aktivitas kehidupan masyarakat Indonesia. Penelitian ini bertujuan untuk mengetahui pola curah hujan menggunakan metode *k-means* dan *k-medoids*. Kedua metode tersebut merupakan metode klasterisasi non-hirarki atau *partial clustering* yang mempartisi data ke dalam satu atau lebih klaster. Perbedaannya terletak pada penentuan pusat klaster atau *centroid*, metode *k-means* menggunakan mean (nilai rataan data) sebagai *centroid* dan metode *k-medoids* menggunakan objek data sebagai perwakilan (*medoids*) sebagai *centroid*. Data yang digunakan adalah data curah hujan satu tahun (12 bulan) di Jakarta tahun 2019. Hasil yang diperoleh yaitu bulan Mei hingga November termasuk klaster satu (musim kemarau). Sementara bulan Januari, Februari, Maret, April, dan Desember termasuk klaster dua (musim penghujan).

Kata Kunci: Curah Hujan, *K-Means*, *K-Medoids*.

1. Pendahuluan

Indonesia merupakan negara yang memiliki dua musim yaitu musim kemarau dan musin penghujan (Anne 2021). Kedua musim tersebut sangat berpengaruh terhadap kelangsungan hidup masyarakat Indonesia. Akibatnya, masyarakat Indonesia harus mengetahui faktor dari kedua musim tersebut agar dapat melakukan aktivitas dengan lancar. Salah satu faktor yang penting dari musim-musim tersebut adalah curah hujan (Pranita 2020).

Curah hujan adalah ketinggian air hujan yang terkumpul di tempat datar yang tidak menyerap, tidak menguap, dan tidak mengalir setelah hujan turun. Pengukuran curah hujan dilakukan dengan satuan tinggi milimeter (mm) di atas permukaan yang datar. Unsur hujan satu milimeter artinya dalam luasan satu meter persegi pada tempat yang datar tertampung air hujan setinggi satu milimeter atau tertampung air hujan sebanyak satu liter (BMKG 2017).

Mengetahui tinggi rendahnya curah hujan dapat dilakukan dengan cara mengelompokkan data curah hujan menjadi beberapa kategori. Salah satu metode yang dapat diterapkan pada kasus ini adalah klasterisasi. Klasterisasi merupakan metode *machine learning* yang melibatkan pengelompokan titik data berdasarkan kesamaan (Larose dan Larose 2015). Klasterisasi termasuk *unsupervised learning method*, yaitu suatu pembelajaran yang menggunakan algoritma *machine learning* untuk menganalisis dan mengelompokkan data yang tidak berlabel. Klasterisasi dibagi menjadi dua tipe yaitu klasterisasi hirarki dan klasterisasi non-hirarki. Klasterisasi non-hirarki disebut juga *partial clustering*, dalam prosesnya semua titik data ditugaskan ke sebuah klaster karena data-data tersebut secara akurat akan merepresentasikan fitur klaster (Joshi 2016).

Salah satu penelitian terdahulu yang berjudul “Sistem Pengelompokan Curah Hujan Menggunakan Metode *K-Means* di Wilayah Kalimantan Timur” membahas tentang klasterisasi curah hujan di suatu wilayah. Penelitian tersebut menggunakan metode *k-means* untuk mengolah data curah hujan satu tahun (12 bulan) pada tahun 2016 dari 10 stasiun pengamatan BMKG Temindung Samarinda. Hasil penelitian tersebut memberikan informasi pengelompokan curah hujan menjadi tiga kategori yaitu rendah, sedang, dan tinggi (Hasanah *et al.* 2017).

Adapun dalam penelitian ini akan dilakukan klasterisasi data curah hujan di Jakarta tahun 2019 yang akan dikategorikan menjadi dua yaitu musim kemarau dan musim penghujan. Penelitian ini bertujuan untuk memberikan informasi tentang pola curah hujan di Jakarta. Algoritma klasterisasi yang akan diterapkan dalam penelitian ini adalah metode *k-means* dan *k-medoids*, keduanya termasuk klasterisasi non-hirarki.

2. Data dan Metode

2.1. Alat dan Bahan

Alat yang digunakan adalah *personal computer* (PC) dengan spesifikasi komputer antara lain *brands* Asus, RAM 4 GB, *storage* 256 G SSD, dan *screen type* 11.6 inch HD. Selanjutnya, *software* yang digunakan yaitu Julia versi 1.6.3. Pada Julia terdapat beberapa metode klasterisasi yang dapat digunakan seperti *k-means* dan *k-medoids*. Semua algoritma tersebut telah tersedia pada paket Clustering.jl. Tambahan pada metode *k-medoids* diperlukan juga paket Distances.jl untuk menghitung jarak Euclid. Adapun sebagai visualisasi hasil memerlukan paket VegaLite.jl.

2.2. Data

Data curah hujan yang digunakan adalah data curah hujan (mm) di Jakarta per jam tahun 2019 dari website <https://www.meteoblue.com/> dengan ukuran-ukuran dasar sebagai berikut.

Tabel 1. Ukuran-ukuran dasar dari data (dalam satuan mm)

Nilai Rata-Rata	Nilai Minimum	Nilai Maksimum
0.181998	0	17.3

Kemudian data per jam tersebut ditotal setiap bulannya, sehingga terdapat 12 data curah hujan yang akan diklasterisasi menggunakan metode *k-means* dan *k-medoids*. Berikut adalah plot data total curah hujan di setiap bulan.



Gambar 1. Plot data total curah hujan setiap bulan

2.3. Metode *K-Means*

K-means merupakan metode klasterisasi non-hirarki yang mempartisi data ke dalam satu atau lebih klaster. Data yang memiliki karakteristik yang sama akan dikelompokkan ke dalam satu klaster yang sama, sehingga metode ini mampu meminimalkan rata-rata jarak setiap data ke klasternya. Lebih lanjut, berikut adalah algoritma metode *k-means* (Joshi 2016).

1. Tentukan k sebagai jumlah klaster yang akan dibentuk.
2. Tentukan k *centroid* awal secara acak.
3. Hitung jarak setiap data ke masing-masing *centroid*. Perhitungan jarak dapat dihitung menggunakan rumus Euclidean Distance, Manhattan, Cosine Similarity, dan lain-lain. Perhitungan jarak menggunakan rumus Euclidean Distance sangat baik untuk menghitung jarak dengan pusat klaster.
4. Alokasikan masing-masing data ke dalam *centroid* terdekat.
5. Tentukan *centroid* baru di masing-masing klaster dengan mencari rata-rata data di setiap klaster.
6. Ulangi langkah 3-5 sampai *centroid* semua klaster tidak lagi berubah.

2.4. Metode *K-Medoids*

Metode *k-medoids* atau sering disebut juga dengan PAM (*Partitioning Around Medoids*) dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw. Metode ini menyerupai metode *k-means* karena kedua metode ini *partitional* yang memecah data menjadi kelompok-kelompok. Perbedaan dari kedua metode tersebut terletak pada penentuan pusat klaster atau *centroid*, metode *k-means* menggunakan mean (nilai rataan data) sebagai *centroid* dan metode *k-medoids* menggunakan objek data sebagai perwakilan (*medoids*) sebagai *centroid*. Algoritma *k-medoids* beroperasi pada prinsip untuk meminimalkan jumlah kesamaan antara setiap objek dan titik referensi yang sesuai. Algoritma *k-medoids* dapat dilakukan dengan Langkah-langkah sebagai berikut (Kaur *et al.* 2014).

1. Tentukan k sebagai jumlah klaster yang akan dibentuk.
2. Tentukan k *medoids* sebagai *centoid* awal secara acak.
3. Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus Euclidean Distance.
4. Alokasikan data-data *non-medoids* ke dalam klaster yang paling dekat dengan *medoids*.
5. Hitung total simpangan (S) dengan menghitung total jarak baru – total jarak lama. Jika $S < 0$ maka tukar objek dengan data klaster *non-medoids* untuk membentuk sekumpulan k objek baru sebagai *medoids*.

6. Ulangi langkah 3-5 hingga tidak terjadi perubahan pada *medoids*, sehingga didapatkan klaster beserta anggota klaster masing-masing.

2.3. Tahapan

Berikut adalah tahapan-tahapan yang dilakukan pada pengerajan penelitian ini.

1. Pengumpulan data.
2. Transformasi data dengan mengakumulasikan data per jam setiap bulan.
3. Pengolahan data menggunakan metode *k-means* dan *k-medoids*.
4. Melakukan visualisasi hasil klasterisasi.

3. Hasil dan Pembahasan

3.1. Metode *k-means*

Berikut adalah Tabel 2 yang menunjukkan hasil klasterisasi data curah hujan menggunakan metode *k-means*.

Tabel 2. Hasil klasterisasi metode *k-means*

Bulan	Total Curah Hujan	Klaster ke-
Januari	198.5	2
Februari	251	2
Maret	290.3	2
April	303.9	2
Mei	83.4	1
Juni	12.5	1
Juli	16.2	1
Agustus	15.1	1
September	2	1
Okttober	31.3	1
November	83.5	1
Desember	306.6	2

Berdasarkan Tabel 2, bulan-bulan yang termasuk ke dalam klaster satu memiliki total curah hujan yang lebih rendah dibandingkan dengan total curah hujan bulan-bulan yang termasuk ke dalam klaster dua. Sehingga dapat digolongkan bahwa klaster satu yaitu bulan Mei hingga bulan November termasuk musim kemarau. Sementara klaster dua yaitu bulan Januari, Februari, Maret, April, dan Desember termasuk musim penghujan.

3.2. Metode *k-medoids*

Berikut adalah Tabel 3 yang menunjukkan hasil klasterisasi data curah hujan menggunakan metode *k-medoids*.

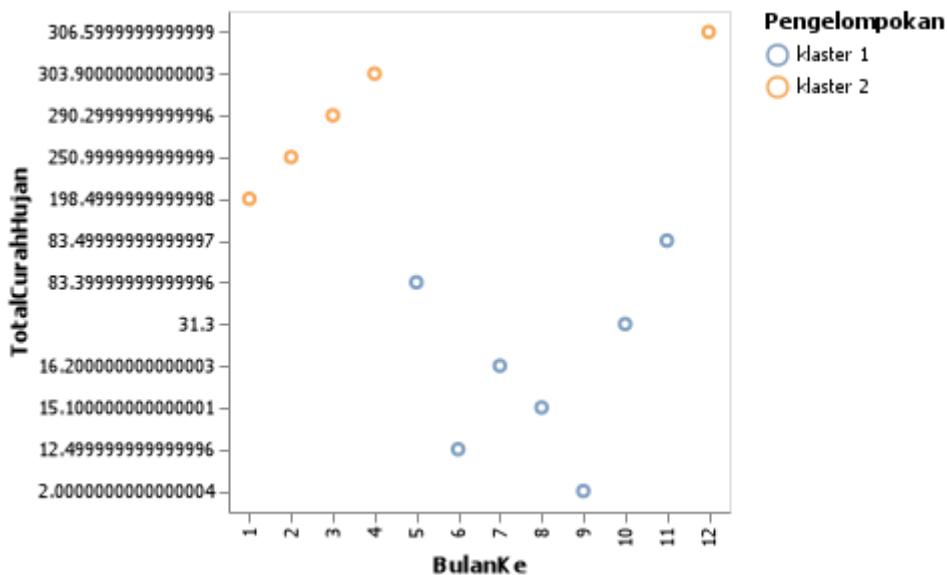
Tabel 3. Hasil klasterisasi metode *k-medoids*

Bulan	Total Curah Hujan	Klaster ke-
Januari	198.5	2
Februari	251	2
Maret	290.3	2
April	303.9	2
Mei	83.4	1
Juni	12.5	1
Juli	16.2	1
Agustus	15.1	1
September	2	1
Okttober	31.3	1
November	83.5	1
Desember	306.6	2

Hasil klasterisasi menggunakan metode *k-medoids* yang disajikan pada Tabel 3 menunjukkan hasil pola yang sama dengan hasil klasterisasi menggunakan metode *k-means* yang disajikan pada Tabel 2. Bulan-bulan pada klaster satu yaitu bulan Mei hingga bulan November termasuk musim kemarau. Sedangkan klaster dua yaitu bulan Januari, Februari, Maret, April, dan Desember termasuk musim penghujan.

3.3. Visualisasi

Berikut adalah gambar visualisasi hasil klasterisasi dari kedua metode yang menunjukkan hasil yang sama.



Gambar 2. Visualisasi hasil klasterisasi

Gambar 2 dengan *point* berwarna biru menunjukkan bulan yang termasuk klaster satu yaitu musim kemarau, dan *point* berwarna oranye menunjukkan bulan yang termasuk klaster dua yaitu musim penghujan.

3.4. Kodingan

Berikut adalah beberapa kode yang digunakan dalam penelitian ini.

Kode 1. Kode untuk melakukan metode *k-means*

```
curah_hujan=Float64.(per_bulan[!, 4])
cluster2 = kmeans(curah_hujan', 2)
insertcols!(per_bulan, 5,:Cluster=>cluster2.assignments)
```

Pendefinisan “curah_hujan” untuk mengambil data curah hujan yang akan diklasterisasi dari tabel “per_bulan” kolom keempat. Kode baris kedua menunjukkan kode untuk proses klasterisasi menggunakan metode *k-means* pada data curah_hujan dengan menentukan banyaknya klaster sebanyak dua. Adapun kode baris ketiga untuk menambahkan kolom hasil klasterisasi pada tabel “per_bulan” di kolom kelima.

Kode 2. Kode untuk melakukan metode *k-medoids*

```
xmatrix = curah_hujan'
D = pairwise(Euclidean()). xmatrix. xmatrix.dims=2
K = kmedoids(D,2)
insertcols!(per_bulan, 6,:medoids_clusters=>K.assignments)
```

Kode baris Ketika menunjukkan kode untuk proses klasterisasi menggunakan metode *k-medoids* pada data curah hujan dengan banyaknya klaster sebanyak dua. Adapun kode baris keempat untuk menambahkan kolom hasil klasterisasi pada tabel “per_bulan” di kolom keenam.

Kode 3. Kode untuk melakukan visualisasi

```
hasil |>
@vlpplot(
    :point,
    x=:BulanKe,
    y={:TotalCurahHujan, sort="descending"},
    color=:Pengelompokan,
    width=250,
    height=250
)
```

Kodingan di atas untuk menampilkan visualisasi hasil klasterisasi data curah hujan dengan perbedaan warna menunjukkan perbedaan klaster.

4. Simpulan

Indonesia merupakan negara yang memiliki dua musim yaitu musim kemarau dan musim penghujan. Salah satu faktor yang penting dari musim-musim tersebut adalah curah hujan. Pengukuran curah hujan dilakukan dengan satuan tinggi milimeter (mm) di atas permukaan yang datar. Unsur hujan satu milimeter artinya dalam luasan satu meter persegi pada tempat yang datar tertampung air hujan setinggi satu milimeter atau tertampung air hujan sebanyak satu liter.

Mengetahui tinggi rendahnya curah hujan dapat dilakukan dengan cara mengelompokkan data curah hujan menjadi beberapa kategori. Salah satu metode yang dapat diterapkan pada kasus ini adalah klasterisasi. Hasil klasterisasi data curah hujan di Jakarta tahun 2019 menggunakan metode *k-means* dan *k-medoids* dengan bantuan *software Julia* versi 1.6.3 diperoleh hasil yang sama. Bulan Mei hingga bulan November termasuk klaster satu. Sedangkan bulan Januari, Februari, Maret, April, dan Desember termasuk klaster dua. Total curah hujan pada bulan-bulan di klaster satu lebih rendah dibandingkan bulan-bulan yang termasuk klaster dua, sehingga dapat dikategorikan bahwa klaster satu termasuk musim kemarau dan klaster dua termasuk musim penghujan.

5. Daftar Pustaka

- Anne. 2021. Pembagian Musim Penghujan dan Kemarau di Indonesia. *Kumparan.com*. [diunduh 2021 Okt 24]. Tersedia pada: <https://kumparan.com/berita-update/pembagian-musim-penghujan-dan-kemarau-di-indonesia-1wEAw0MzdzQ>
- [BMKG] Badan Meteorologi Klimatologi dan Geofisika. 2017. Daftar Istilah Klimatologi. [diunduh 2021 Okt 9]. Tersedia pada: <http://balai3.denpasar.bmkg.go.id/>
- Hasanah N, Ugiarto M, Puspitasari N. 2017. Sistem pengelompokan curah hujan menggunakan metode k-means di wilayah Kalimantan Timur. *Pros. Semin. Nas. Ilmu Komput. Dan Teknologi Inf.* 2(2):6.
- Joshi A. 2016. *Julia for Data Science: Explore the World of Data Science from Scratch with Julia by Your Side*. Birmingham: Packt.
- Kaur NK, Kaur U, Singh D. 2014. K-Medoid Clustering Algorithm - A Review. *Int. J. Comput. Appl. Technol.* 1(1):42–45.
- Larose DT, Larose CD. 2015. *Data Mining and Predictive Analytics*. Second. US: Wiley. Ed ke-Google-Books-ID: 0HJuBwAAQBAJ.
- Pranita E. 2020. Musim Hujan, Ketahui 4 Faktor Pemicu Banjir dan Mitigasinya. *KOMPAS.com*. [diunduh 2021 Okt 24]. Tersedia pada: <https://www.kompas.com/sains/read/2020/12/04/170600123/musim-hujan-ketahui-4-faktor-pemicu-banjir-dan-mitigasinya>

Analisis Pengaruh Suhu Udara dan Curah Hujan Terhadap Kelembaban Udara di Jakarta pada Tahun 2019

Trianty Putri Blante

Departemen Matematika, IPB University, Bogor 16680, Indonesia. email: putriblante@apps.ipb.ac.id

Abstrak.

Indonesia merupakan negara beriklim tropis dengan suhu dan kelembaban yang tinggi. Kelembaban udara adalah banyaknya uap air yang terkandung dalam udara atau atmosfer. Suhu udara adalah keadaan panas atau dinginnya udara atau juga bisa disebut dengan temperatur. Variasi suhu udara di kepulauan Indonesia tergantung pada ketinggian tempat. Suhu udara akan semakin rendah pada tempat yang semakin tinggi, apabila suhu suatu tempat tinggi, maka kelembabannya rendah dan sebaliknya, apabila suhu rendah, maka kelembaban tinggi. Curah hujan adalah ketinggian air hujan dalam tempat yang datar, tidak menguap, tidak meresap dan tidak mengalir. Tujuan penelitian ini adalah untuk membahas seberapa besar pengaruh suhu udara terhadap kelembaban udara dan pengaruh curah hujan terhadap kelembaban udara di Jakarta per minggu pada tahun 2019. Berdasarkan analisis diperoleh nilai korelasi pengaruh suhu udara terhadap kelembaban udara yaitu sebesar 0,5106 dari hasil pendekatan menggunakan metode Regresi Linear Sederhana. Dapat disimpulkan dalam penggunaan metode Regresi Linear Sederhana dengan model $Y = 248,765 - 6,14925X + \varepsilon$. Dan nilai korelasi pengaruh curah hujan terhadap kelembaban udara yaitu sebesar 0,8033 dari hasil pendekatan menggunakan metode Regresi Linear Sederhana dengan model $Y = 72,9306 + 0,137738X + \varepsilon$.

Kata Kunci: Suhu Udara, Kelembaban Udara, Curah Hujan, Regresi Linear.

1. Pendahuluan

Indonesia dikenal sebagai bagian dari benua maritim dengan ciri khasnya berupa daratan yang dikelilingi laut, di mana aktivitas konveksi yang terjadi merupakan bagian dari sirkulasi global (Nuryanto, 2012). Wilayah Indonesia berada di antara 6°LU-11°LS dan 95°BT-141°BT dan merupakan daerah tropis dengan dua musim, yaitu: musim kemarau dan musim penghujan. Berdasarkan klasifikasi iklim global, wilayah kepulauan Indonesia sebagian besar tergolong dalam zona iklim tropis basah dan sisanya masuk zona iklim pegunungan (Gumilanggeng, 2013).

Kelembaban udara adalah banyaknya uap air yang terkandung dalam udara atau atmosfer (Swarinoto & Sugiyono, 2011). Kandungan uap air dalam udara hangat lebih banyak daripada kandungan uap air di dalam udara dingin. Jika kandungan uap air di udara mengalami pendinginan, maka akan terbentuk titik-titik air. Titik-titik air atau biasa disebut dengan uap air adalah suatu gas, yang tidak dapat dilihat, yang merupakan salah satu bagian dari atmosfer. Banyaknya uap air yang dikandung, tergantung pada suhu udara. Semakin tinggi suhu udara, makin banyak uap air yang terkandung (Hardjodinomo, 1975).

Suhu udara adalah keadaan panas atau dinginnya udara atau juga bisa disebut dengan temperatur (Siswanti, 2011). Variasi suhu udara di kepulauan Indonesia tergantung pada ketinggian tempat. Suhu udara akan semakin rendah pada tempat yang semakin tinggi (Lakitan, 1994). Apabila suhu suatu tempat tinggi, maka kelembabannya rendah dan sebaliknya, apabila suhu rendah, maka kelembaban tinggi. Di mana hal ini suhu dan kelembaban berkaitan. Suhu udara dan kelembaban udara memiliki kandungan air. Suhu udara yang tinggi juga mengembangkan volume udara sehingga membuat tekanan udara menjadi rendah.

Curah hujan adalah ketinggian air hujan dalam tempat yang datar, tidak menguap, tidak meresap dan tidak mengalir. Curah hujan 1 mm, artinya dalam luasan satu meter persegi terdapat tempat yang datar tertampung air setinggi satu milimeter atau tertampung air sebanyak satu liter dalam jangka waktu tertentu. Curah hujan dapat berupa butir-butir air atau kristal es yang jatuh atau keluar dari awan (Swarinoto & Sugiyono, 2011). Menurut Kamus Besar Bahasa Indonesia (KBBI, 2012), curah hujan adalah banyaknya hujan yang turun di suatu daerah pada jangka waktu tertentu.

Analisis regresi merupakan suatu cara yang dapat digunakan untuk mengetahui hubungan sebuah variabel tak bebas (terikat) dengan satu atau lebih variabel bebas. Analisis regresi dapat digunakan untuk menganalisis data dan mengambil kesimpulan tentang suatu hubungan ketergantungan variabel terhadap variabel lainnya. Berdasarkan jumlah variabel bebas, analisis regresi dapat dibagi dua macam yaitu, analisis regresi linear sederhana dan analisis regresi linear berganda.

Adanya metode analisis regresi ini sangat menguntungkan bagi banyak pihak, baik dibidang sains, industri maupun bisnis. Salah satu manfaat analisis regresi adalah memperkirakan suatu kejadian yang akan terjadi dengan menganalisis penyebab yang mungkin mempengaruhi kejadian tersebut. Oleh karena itu makalah ini akan membahas seberapa besar pengaruh suhu udara terhadap kelembaban pada bulan Januari - Desember di Jakarta tahun 2019 dan seberapa besar

pengaruh curah hujan terhadap kelembaban pada bulan Januari - Desember di Jakarta tahun 2019 sehingga mendapatkan pemodelan yang tepat tentang pengaruh suhu udara dan curah hujan terhadap kelembaban udara.

2. Data dan Metode

Data yang digunakan adalah data cuaca (suhu, curah hujan, kelembaban, kecepatan dan arah angin) di Jakarta per jam tahun pada 2019. Namun, data yang dianalisis pada makalah ini adalah data cuaca (suhu, curah hujan dan kelembaban) di Jakarta per minggu tahun 2019. Metode yang digunakan adalah metode Regresi Linear sederhana dan Analisis Korelasi Pearson.

2.1. Alat dan Bahan

Penelitian ini menggunakan *personal computer* (PC) dengan spesifikasi:

<i>Operating System</i>	Windows 10 Home Single Language 64-bit
<i>Processor</i>	Intel(R) Core(TM) i7-8565U
<i>Memory</i>	16384MB RAM

Penelitian ini menggunakan *Microsoft Excel* untuk mengolah data dan menggunakan *software Julia* versi 1.6.2. dengan paket sebagai berikut:

<i>XLSX</i>	Untuk membaca dan menulis <i>file spreadsheet Excel</i> .
<i>DataFrame</i>	Untuk menyediakan seperangkat alat untuk berkerja dengan data.
<i>Plots</i>	Untuk <i>Plotting</i> .
<i>GLM</i>	Untuk menyesuaikan model linear untuk data yang diberikan.
<i>Statistics</i>	bekerja dengan data dan melakukan <i>statistic</i> di <i>Julia</i> .

2.2. Data

Sumber data diperoleh dari data cuaca di Jakarta per jam tahun 2019. ([website: *https://www.meteoblue.com/*](https://www.meteoblue.com/)). Jenis data adalah data sekunder. Data sekunder adalah data yang sudah diolah terlebih dahulu dan baru didapatkan oleh peneliti dari sumber yang lain sebagai tambahan informasi.

2.3. Metode Regresi Linear Sederhana

Regresi Linear Sederhana adalah metode statistik yang berfungsi untuk menguji sejauh mana hubungan sebab akibat antara variabel faktor penyebab (X) terhadap variabel akibatnya. Setelah harga a dan b ditemukan, maka persamaan regresi linear sederhana dapat disusun. Model persamaan regresi linear sederhana adalah seperti berikut ini:

$$Y = a + bX + \varepsilon$$

dengan,

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n \sum X^2 - (\sum X)^2}$$

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2}$$

Keterangan :

Y = Variabel response atau variabel akibat (*dependent*)

X = Variabel predictor atau variabel faktor penyebab (*Independent*)

a = Konstanta

b = Koefisien regresi (kemiringan); besaran *response* yang ditimbulkan oleh *predictor*.

ε = error

Persamaan regresi yang telah ditemukan itu dapat digunakan untuk melakukan prediksi (ramalan) bagaimana dalam variabel dependen akan terjadi bila individu dalam variabel independen ditetapkan.

Pengambilan harga-harga X untuk meramalkan Y harus dipertimbangkan secara rasional dan menurut pengalaman, yang masih berada pada batas ruang gerak X (Ismada, 2012).

2.4. Analisis Korelasi Pearson

Analisis Korelasi adalah metode statistika yang digunakan untuk menentukan suatu besaran yang menyatakan bagaimana kuat hubungan suatu variabel dengan variabel lain dengan tidak mempersoalkan apakah suatu variabel tertentu tergantung kepada variabel lain (Sekaran, 2010). Hubungan yang dimiliki dua variabel atau lebih untuk

mengukur kekuatan hubungan antara satu variabel dengan variabel lainnya, untuk mengetahui apakah ada atau tidaknya hubungan antara satu variabel dengan variabel lainnya (Syafrizal Helmi, 2010).

Analisis korelasi *Pearson* digunakan untuk mengetahui hubungan antara satu variabel dengan variabel yang lain secara linear. Data yang digunakan berskala interval atau rasio. Nilai korelasi r adalah 0 sampai 1, semakin mendekati 1 hubungan yang terjadi semakin kuat. Sebaliknya, nilai semakin mendekati 0 maka hubungan yang terjadi semakin lemah.

Menurut Sugiyono (2010) pedoman untuk menginterpretasikan hasil koefisien korelasi sebagai berikut :

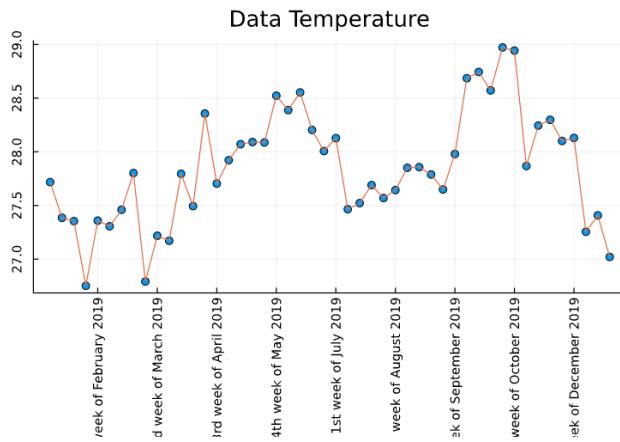
- 0,00 - 0,199 : sangat rendah
- 0,20 - 0,399 : rendah
- 0,40 - 0,599 : sedang
- 0,60 - 0,799 : kuat
- 0,80 - 1,000 : sangat kuat

Berdasarkan pedoman di atas data yang semakin mendekati 1 maka data tersebut semakin valid.

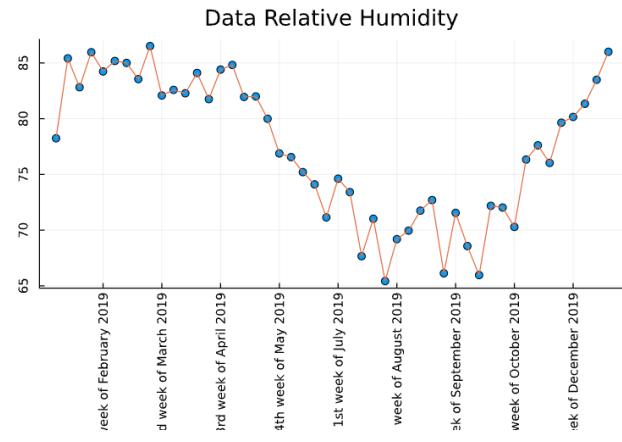
Koefisien korelasi adalah ukuran yang dipakai untuk mengetahui derajat hubungan antara variabel-variabel (Siregar, 2013). Nilai koefisien korelasi berada dia antara $-1 < r < 1$ yaitu apabila $r=-1$ korelasi negatif sempurna, artinya taraf signifikansi dari pengaruh variabel X terhadap variabel Y sangat lemah, dan apabila $r=1$ korelasi positif sempurna, artinya taraf signifikansi dari pengaruh variabel X terhadap variabel Y sangat kuat (Sudjana, 2005).

3. Hasil dan Pembahasan

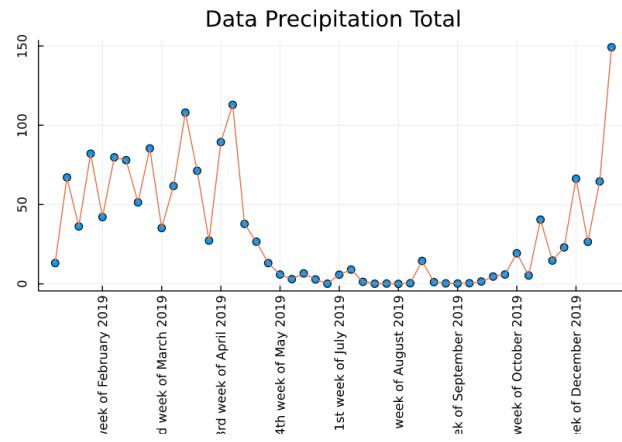
Data yang digunakan adalah data cuaca (suhu, curah hujan dan kelembaban) di Jakarta per minggu pada tahun 2019. penulis akan melakukan Analisis Regresi Sederhana menggunakan Julia dan penulis akan melakukan prediksi terhadap rata-rata kelembaban udara, rata-rata suhu udara telah dan jumlah tingkat curah hujan.



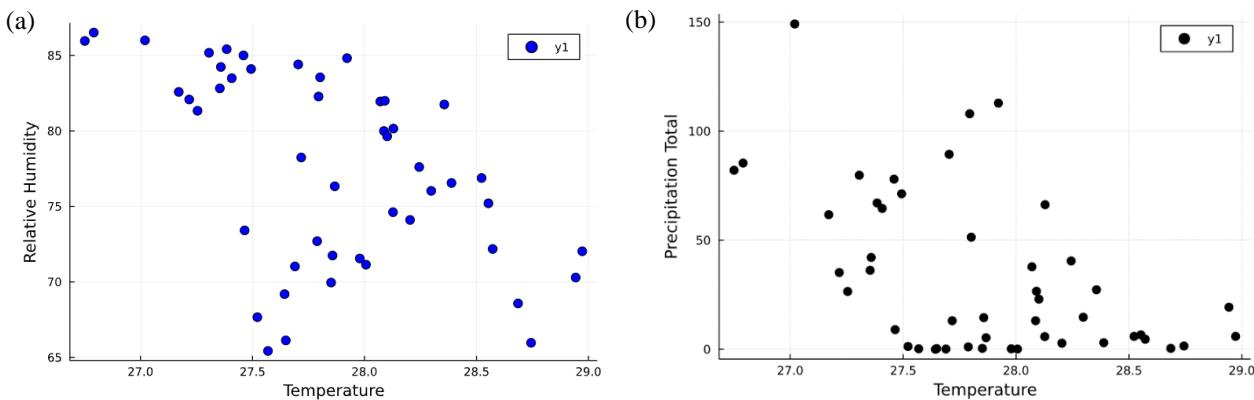
Gambar 1. *Temperature* per minggu di Jakarta tahun 2019



Gambar 2. *Relative humidity* per minggu di Jakarta tahun 2019

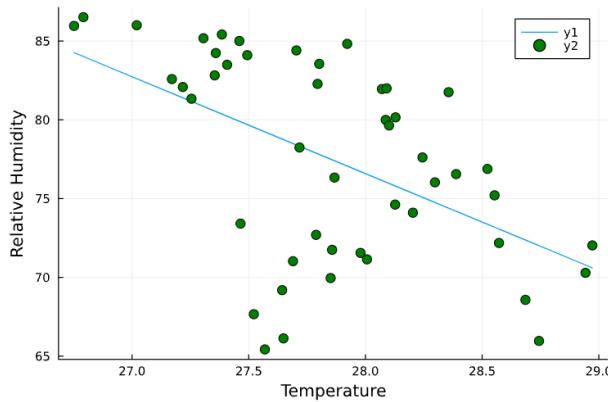


Gambar 3. *Precipitation total* per minggu di Jakarta tahun 2019

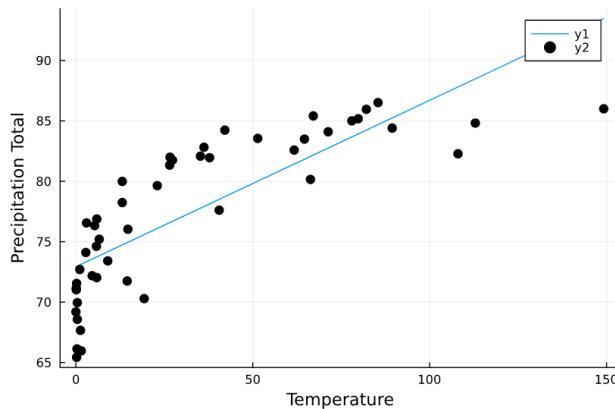


Gambar 4. (a) *Plot* penyebaran dari data *Temperature* dan *Relative Humidity* (b) *Plot* penyebaran dari data *Precipitation Total* dan *Relative Humidity*

Dari data di atas dapat diperoleh sebaran data dari *Temperature* dan *Relative Humidity* serta sebaran data dari *Precipitation Total* dan *Relative Humidity* dan penulis akan menganalisis menggunakan metode Regresi Linear sederhana. Selanjutnya, penulis akan membuat model Regresi Linear sederhana menggunakan *software Julia* kemudian diperoleh:



Gambar 5. Regresi Linear sederhana pengaruh *Temperature* terhadap *Relative Humidity*



Gambar 6. Regresi Linear sederhana pengaruh *Precipitation Total* terhadap *Relative Humidity*

Variabel X mendefinisikan rata-rata *Temperature* (suhu) dan *Precipitation Total* (curah hujan) yang merupakan variabel *Independent* atau variabel yang mempengaruhi atau variabel bebas (tidak ada yang mempengaruhi nilainya), dan variabel Y mendefinisikan rata-rata *Relative Humidity* (kelembaban) yang merupakan variabel *dependent* atau variabel yang dipengaruhi atau variabel tidak bebas (ada yang mempengaruhi nilainya).

Kemudian penulis akan mencari *output* model yang terbentuk untuk mengetahui nilai dari koefisien a dan b menggunakan *software Julia* dan diperoleh :

$$Q \sim 1 + P$$

Coeficients :

Tabel 1. Output model yang terbentuk dari pengaruh *Temperature* terhadap *Relative Humidity*

	Coef.	Std. Error	t	Pr(> t)	Lower 95%	Upper 95%
(Intercept)	248.765	42.5327	5.85	<1e-06	163.151	334.379
P	-6.14925	1.52684	-4.03	0.0002	-9.22263	-3.07588

Pada tabel di atas terdapat tanda “~” maksud yaitu nilai Y dipengaruhi oleh nilai X dimakalah ini Y dilambangkan dengan Q terhadap X yang dilambangkan dengan P dan diperoleh model $Y = 248,765 - 6,14925X + \varepsilon$

Kemudian, penulis akan menghitung *R square* dan korelasi *pearson* menggunakan *software Julia* yaitu *R square* dari analisis pengaruh suhu udara terhadap kelembaban udara adalah 0.2607 atau 26%. Dan korelasi *pearson* yang diperoleh antara variabel *Temperature* dan *Relative Humidity* adalah 0.5106.

Selanjutnya, penulis akan mencari *output* model yang terbentuk untuk mengetahui nilai dari koefisien a dan b menggunakan *software Julia* dan diperoleh :

$$Q \sim 1 + R$$

Coeficients :

Tabel 2. Output model yang terbentuk dari pengaruh *Precipitation Total* terhadap *Relative Humidity*

	Coef.	Std. Error	t	Pr(> t)	Lower 95%	Upper 95%
(Intercept)	72.9306	0.744301	97.99	<1e-54	71.4324	74.4288
P	0.137738	0.0150583	9.15	<1e-11	0.107427	0.168049

Dapat dilihat pada table diatas terdapat tanda “~” maksud yaitu nilai Y dipengaruhi oleh nilai X dimakalah ini Y dilambangkan dengan Q terhadap X yang dilambangkan dengan R dan diperoleh model $Y = 72,9306 + 0,137738X + \varepsilon$

Kemudian, penulis akan menghitung *R square* dan korelasi *pearson* menggunakan *software Julia* dan diperoleh *R square* dari analisis pengaruh curah hujan terhadap kelembaban udara adalah 0.6452 atau 64%. Dan korelasi *pearson* yang diperoleh antara variabel *Temperature* dan *Relative Humidity* adalah 0.8033.

Dari hasil laporan yang telah dilakukan di atas, maka didapat beberapa hal sebagai berikut:

1. Didapatkan *p-value* dari kedua data < 0.05 yang artinya terdapat pengaruh nilai X terhadap y.
2. Dari hasil analisis di atas diperoleh model pengaruh suhu udara terhadap kelembaban udara yaitu

$$Y = 248,765 - 6,14925X + \varepsilon$$

model tersebut mampu menjelaskan sebesar 26% dan data antar pengamatan saling berkorelasi.

3. Dari hasil analisis di atas diperoleh model pengaruh suhu udara terhadap kelembaban udara yaitu

$$Y = 72,9306 + 0,137738X + \varepsilon$$

model tersebut mampu menjelaskan sebesar 64% dan data antar pengamatan saling berkorelasi.

4. Korelasi *pearson* antara variabel *Temperature* dan *Relative Humidity* sebesar 0,5106. Karena nilai korelasi berada di range 0,40 - 0,599 maka disimpulkan bahwa hubungan antara kedua variabel tersebut sedang.
5. Korelasi *pearson* antara variabel *Precipitation Total* dan *Relative Humidity* sebesar 0,8033. Karena nilai korelasi berada di range 0,80 - 1,000 maka disimpulkan bahwa hubungan antara kedua variabel tersebut sangat kuat.

4. Simpulan

Dapat disimpulkan bahwa terdapat pengaruh *Precipitation Total* terhadap *Relative Humidity* yang lebih baik dari pengaruh *Temperature* terhadap *Relative Humidity* dan diperoleh model $Y = 72,9306 + 0,137738X + \varepsilon$, model tersebut mampu menjelaskan sebesar 64% dan nilai korelasi $r=0,8033$ yang menunjukkan hubungan sangat kuat antara kedua variabel tersebut.

5. Ucapan Terima Kasih

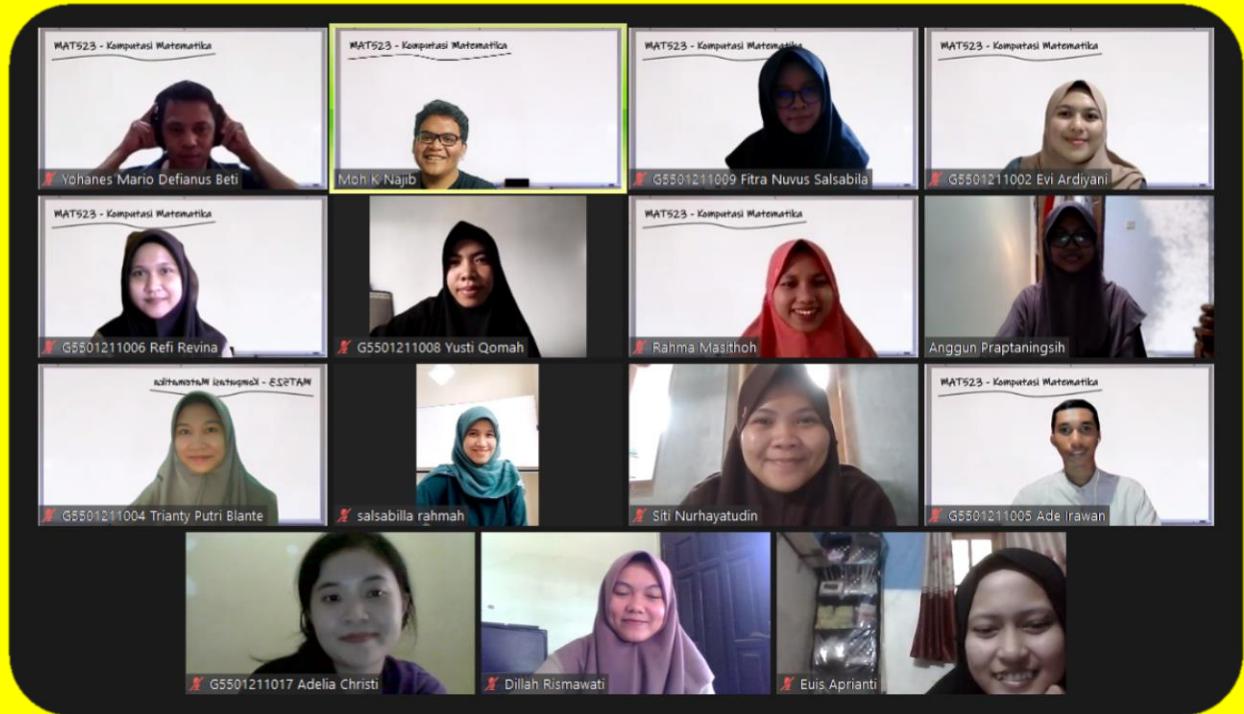
Dalam penulisan makalah ini, penulis mendapat banyak ide, saran, kritik serta semangat dari beberapa pihak dengan harapan agar makalah ini lebih baik. Oleh karena itu penulis ingin mengucapkan terima kasih kepada Refi Revina, Evi Ardiyani, Anggun Praptaningsih, dan Fitra Nuvus Salsabila atas bantuannya selama proses pembuatan makalah ini.

6. Daftar Pustaka

Gumilanggeng, E. 2013. *Visualiasi Informasi Klasifikasi Iklim Koppen Menggunakan Metode Polygon Thiessen*.

Skripsi. Salatiga: FTI Universitas Kristen Satya Wacana.

- Hardjodinomo, S. 1975. Ilmu Iklim dan Pengairan. Bandung: Binacipta.
- Helmi, S. 2010. *Analisis Data*. Medan : USU Press.
- Ismada, R. 2012. ANALISIS REGRESI. Diambil kembali dari Wordpress.com:
<https://rufiismada.files.wordpress.com/2012/02/analisis-regresi.pdf>
- Kho. D. (t.thn). *Analisis Regresi Linear Sederhana (Simple Linear Regression)*. Diambil kembali dari Teknik Elektronika: <https://teknikelektronika.com/analisis-regresi-linear-sederhana-simple-linear-regression/>
- Lakitan, B. 1994. Dasar-Dasar Klimatologi. Jakarta: PT. Raja Grafindo Persada.
- Sekaran, Uma dan Bougie, R., 2010, *Research Methods for Business: A SkillBuilding Approach*, John Wiley and sons, inc. : London.
- Siregar, Syofian. 2013. *Statistik Parametrik untuk Penelitian Kualitatif*. Jakarta: Bumi Aksara.
- Siswanti, K. Y. 2011. Model Fungsi Transfer Multivariat dan Aplikasinya untuk meramalkan Curah Hujan Di Kota Yogyakarta. Skripsi. Yogyakarta: FMIPA Universitas Negeri Yogyakarta.
- Sudjana. 2005. *Metoda Statistika*. Tarsito, Bandung.
- Swarinoto, Y. S., Sugiyono. 2011. Pemanfaatan Suhu Udara dan Kelembaban Udara dalam Persamaan Regresi untuk Simulasi Prediksi Total Hujan Bulanan di Bandar Lampung. *Jurnal Meteorologi dan Geofisika*, 12(3):271-281.



”Tuntutlah Ilmu! Ketika kamu miskin, ia akan menjadi hartamu.

Ketika kamu kaya, ia akan menjadi perhiasanmu.”

(Luqman al-Hakim)