

Jurnal Ilmiah Teori dan Aplikasi Statistika

- ❑ **Peramalan Data Ekspor Nonmigas Provinsi Kalimantan Timur Menggunakan Metode Weighted Fuzzy Time Series Lee**
Muhammad Wahdeni Pramana¹, Ika Purnamasari², Surya Prangga³
Universitas Mulawarman
- ❑ **Peramalan Indeks Harga Saham PT Verena Multi Finance Tbk Dengan Metode Pemodelan ARIMA Dan ARCH-GARCH**
Fajriyah Fitriyani¹, Salsabila Fasya A², Muhammad Irfan R³, Teguh Ammar T⁴
Universitas Padjadjaran
- ❑ **Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Provinsi Jawa Barat dengan Pendekatan Regresi Nonparametrik Spline Truncated**
Andrea Tri Rian Dani¹, Ludia Ni'matuzzahroh²
Institut Teknologi Sepuluh Nopember
- ❑ **Analisis Pengaruh Kepadatan Penduduk Dan Jumlah Lansia Terhadap Kasus Kematian Akibat Covid-19 di Provinsi DKI Jakarta Menggunakan Multiple Classification Analysis (MCA)**
Habni Hamara Azmaty¹, Hussein Adi Permana², Lisa Agustina³, Muhamad Fikri Ramdhani⁴, Naufal Abdul Rafi Zaqi⁵, Risni Julaeni Yuhan⁶
Politeknik Statistika STIS
- ❑ **Membandingkan Seleksi Variabel Pada Data Microarray Menggunakan Important Variable Value dan Genetic Algorithm (Studi Kasus Lung Cancer Dataset dan Prostate Cancer Dataset)**
Diana Nurlaili¹, Farida Nur Hayati², Elly Pusporani³
Institut Teknologi Kalimantan
- ❑ **Pemodelan Regresi Spatial Autoregressive Fixed Effect Model Data Panel Pada Tingkat Kemiskinan Di Provinsi Jawa Barat**
Muhammad Irfan R¹, Teguh Ammar T²
Universitas Padjadjaran
- ❑ **Analisis Sentimen Netizen Twitter terhadap Pemberitaan PPN Sembako dan Jasa Pendidikan dengan Pendekatan Social Network Analysis dan Naive Bayes Classifier**
Joko Ade Nursiyono¹, Chusnul Chotimah²
Badan Pusat Statistik Provinsi Jawa Timur

Penanggung Jawab	
Alfisyahrina Hapsery, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Pimpinan Redaksi	
Muhammad Athoillah, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Editor Pelaksana	
Sari Cahyaningtiyas, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Fenny Fitriani, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Nur Silviyah Rahmi, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Sekar Utami Wijaya S.Stat., M.Si	(Universitas PGRI Adi Buana Surabaya)
Mitra Bestari (Reviewer)	
Dr.rer.pol. Dedy Dwi Prastyo, M.Si	(Institut Teknologi Sepuluh Nopember)
Dr. Drs. Agus Suharsono, M.S	(Institut Teknologi Sepuluh Nopember)
Dr. Bambang Widjanarko Otok	(Institut Teknologi Sepuluh Nopember)
Novri Suhermi, S.Si., M.Si., M.Sc	(Institut Teknologi Sepuluh Nopember)
Shofi Andari, S.Stat., M.Si	(Institut Teknologi Sepuluh Nopember)
Dr. RB Fajriya Hakim, S.Si., M.Si	(Universitas Islam Indonesia)
A'yunin Sofro, S.Si., M.Si., Ph.D.	(Universitas Negeri Surabaya)
Arief Rachman Hakim, S.Si., M.Si	(Universitas Diponegoro)
Dani Al Mahkya, S.Si., M.Si	(Sains Aktuaria Institut Teknologi Sumatera)
Dr. Sri Harini	(Universitas Islam Negeri Maulana Malik Ibrahim)
Winda Aprianti, S.Si., M.Si	(Politeknik Negeri Tanah Laut)
Dr. Faula Arina, M.Si	(Universitas Sultan Agung Tirtayasa)
Artanti Indrasetimingsih, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Gangga Anuraga, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Edy Sulisyawan, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)

J Statistika: **Jurnal Ilmiah Teori dan Aplikasi Statistika** merupakan media penerbitan dan pembahasan karya ilmiah dalam bidang ilmu statistika beserta aplikasinya, baik berupa hasil penelitian, bahasan tentang teori, metodologi, komputasi maupun aplikasi statistika dalam bidang lain. Jurnal ini akan terbit dua kali dalam setahun setiap bulan Juli dan Desember.

Redaksi mengundang para pakar dan praktisi dari dalam dan luar lingkungan Universitas PGRI Adi Buana Surabaya untuk menuliskan karya ilmiahnya yang relevan dengan bidang ilmu statistika. Naskah dikirim dengan format A4 margin 4-3-3-3 times new roman 10 dan 11, spasi 1, disertai biodata penulis ke email redaksi: jstat@unipasby.ac.id. Cp: 085645358346

Pengantar Redaksi

Syukur Alhamdulillah bahwa Jurnal Ilmiah: JStatistika, Program Studi Statistika Universitas PGRI Adi Buana Surabaya volume 14 nomor 1 dapat terbit bulan Juli 2021 ini. Berbagai hambatan dapat kita atasi, semoga hambatan-hambatan tersebut tidak akan terjadi lagi pada penerbitan selanjutnya. Jurnal Ilmiah JStatistika terbitan kali ini memuat artikel yang membahas tentang **“Peramalan Data Ekspor Nonmigas Provinsi Kalimantan Timur Menggunakan Metode Weighted Fuzzy Time Series Lee; Peramalan Indeks Harga Saham PT Verena Multi Finance Tbk Dengan Metode Pemodelan ARIMA Dan ARCH-GARCH; Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Provinsi Jawa Barat dengan Pendekatan Regresi Nonparametrik Spline Truncated; Analisis Pengaruh Kepadatan Penduduk Dan Jumlah Lansia Terhadap Kasus Kematian Akibat Covid-19 di Provinsi DKI Jakarta Menggunakan Multiple Classification Analysis (MCA); Membandingkan Seleksi Variabel Pada Data Microarray Menggunakan Important Variable Value dan Genetic Algorithm (Studi Kasus Lung Cancer Dataset dan Prostate Cancer Dataset); Pemodelan Regresi Spatial Autoregressive Fixed Effect Model Data Panel Pada Tingkat Kemiskinan Di Provinsi Jawa Barat; Analisis Sentimen Netizen Twitter terhadap Pemberitaan PPN Sembako dan Jasa Pendidikan dengan Pendekatan Social Network Analysis dan Naive Bayes Classifier”**.

Jurnal Ilmiah Statistika menerima artikel ilmiah dari hasil penelitian, laporan / studi kasus, kajian/ tinjauan pustaka, maupun penyegar ilmu statistika, yang berorientasi pada kemuthakhiran ilmu pengetahuan statistika agar dapat menjadi sumber informasi ilmiah yang mampu memberikan kontribusi dalam mengatasi permasalahan yang semakin kompleks. Redaksi mengundang berbagai ilmuwan dari berbagai lembaga pendidikan tinggi maupun penelitian untuk memberikan sumbangan ilmiahnya, baik berupa hasil penelitian maupun kajian ilmiah dalam bidang statistika dan aplikasinya.

Redaksi sangat mengharapkan masukan-masukan dari para pembaca, professional bidang statistika, atau yang terkait dengan penerbitan, demi semakin meningkatnya kualitas jurnal sebagaimana harapan kita bersama. Semoga artikel ilmiah yang termuat dalam Jurnal Ilmiah JStatistika bermanfaat bagi para akademisi dan professional yang berkecimpung dalam dunia statistika dan aplikasinya.

Surabaya, Juli 2021

Pimpinan Redaksi

DAFTAR ISI

HALAMAN JUDUL

PENGANTAR REDAKSI

DAFTAR ISI

Peramalan Data Ekspor Nonmigas Provinsi Kalimantan Timur Menggunakan Metode Weighted Fuzzy Time Series Lee

Muhammad Wahdeni Pramana¹, Ika Purnamasari², Surya Prangga³

Universitas Mulawarman

1-10

Peramalan Indeks Harga Saham PT Verena Multi Finance Tbk Dengan Metode Pemodelan ARIMA Dan ARCH-GARCH

Fajriyah Fitriyani¹, Salsabila Fasya A², Muhammad Irfan R³, Teguh Ammar T⁴

Universitas Padjadjaran

11-23

Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Provinsi Jawa Barat dengan Pendekatan Regresi Nonparametrik Spline Truncated

Andrea Tri Rian Dani¹, Ludia Ni'matuzzahroh²

Institut Teknologi Sepuluh Nopember

24-29

Analisis Pengaruh Kepadatan Penduduk Dan Jumlah Lansia Terhadap Kasus Kematian Akibat Covid-19 di Provinsi DKI Jakarta Menggunakan Multiple Classification Analysis (MCA)

Habni Hamara Azmaty¹, Hussein Adi Permana², Lisa Agustina³, Muhamad Fikri Ramdhani⁴, Naufal Abdul Rafi Zaqi⁵, Risni Julaeli Yuhan⁶

Politeknik Statistika STIS

30-37

Membandingkan Seleksi Variabel Pada Data Microarray Menggunakan Important Variable Value dan Genetic Algorithm (Studi Kasus Lung Cancer Dataset dan Prostate Cancer Dataset)

Diana Nurlaily¹, Farida Nur Hayati², Elly Pusporani³

Institut Teknologi Kalimantan

38-43

Pemodelan Regresi Spatial Autoregressive Fixed Effect Model Data Panel Pada Tingkat Kemiskinan Di Provinsi Jawa Barat

Muhammad Irfan R¹, Teguh Ammar T²

Universitas Padjadjaran

44-51

Analisis Sentimen Netizen Twitter terhadap Pemberitaan PPN Sembako dan Jasa Pendidikan dengan Pendekatan Social Network Analysis dan Naive Bayes Classifier

Joko Ade Nursiyono¹, Chusnul Chotimah²

Badan Pusat Statistik Provinsi Jawa Timur

52-58

Peramalan Data Ekspor Nonmigas Provinsi Kalimantan Timur menggunakan Metode Weighted Fuzzy Time Series Lee

Muhammad Wahdeni Pramana⁽¹⁾, Ika Purnamasari⁽²⁾, Surya Prangga⁽³⁾

Universitas Mulawarman

Jl. Kuaro, Gn. Kelua, Kota Samarinda, Kalimantan Timur

e-mail: wahdeni.pramana@gmail.com

ABSTRAK

Ekspor merupakan aktivitas perdagangan atau penjualan barang dari dalam negeri ke luar negeri. Ekspor nonmigas sebagai salah satu komponen pembentuk Produk Domestik Regional Bruto (PDRB) sehingga perlu adanya suatu peramalan nilai di masa mendatang. *Fuzzy Time Series* (FTS) merupakan metode peramalan dengan berdasarkan teori himpunan *fuzzy*, logika *fuzzy*, serta hasil peramalan yang dapat dibahasakan (linguistik). Metode *Weighted Fuzzy Time Series* (WFTS) Lee merupakan perluasan dari metode FTS dengan penambahan pembobotan pada tiap pola relasi yang terbentuk. Tujuan penelitian ini adalah memperoleh nilai peramalan ekspor nonmigas Provinsi Kalimantan Timur pada bulan November 2020 serta memperoleh nilai akurasi peramalan berdasarkan metode *Mean Absolute Percentage Error* (MAPE) dan *Root Mean Square Error* (RMSE). Berdasarkan hasil analisis diperoleh nilai akurasi peramalan untuk data Ekspor Nonmigas Provinsi Kalimantan Timur bulan Januari 2019 – Oktober 2020 dengan konstanta pembobot (c) $1 \leq c \leq 2$ menggunakan metode MAPE diperoleh hasil keseluruhan dibawah 10% sehingga diperoleh konstanta pembobot terbaik yaitu $c = 1$ dengan nilai MAPE terminimum yaitu sebesar 3,62% dan RMSE minimum sebesar 50,67. Dari hasil tersebut, diperoleh hasil peramalan untuk bulan November 2020 dengan menggunakan konstanta pembobot terbaik $c = 1$ yaitu sebesar 850,96 juta USD.

Kata kunci: Ekspor, MAPE, nonmigas, RMSE, WFTS Lee.

ABSTRACT

Exports is a trading and sales activity from domestic to overseas. Exports non-oil and gas as a one of component former gross regional domestic product so that its necessary to forecast a value in the future. Fuzzy Time Series (FTS) is a forecasting method based on fuzzy set theory, fuzzy logic, and the result in the form of linguistic. Weighted Fuzzy Time Series (WFTS) Lee was developed from FTS method with additions weighting in each pattern relation that is formed. This paper proposed to forecast value non-oil and gas exports in East Kalimantan Province in November 2020 and to obtain accuracy values using Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) methods. The results of this paper that obtained forecast accuracy values for data exports non-oil and gas East Kalimantan Province from Januari 2019 – Oktober 2020 with constant weighter (c) $1 \leq c \leq 2$ using MAPE as a whole below 10% so that its obtained optimum constant weighter is $c = 1$ with MAPE minimum value 3,62% and RMSE minimum value 50,67. Of these results, the forecast value for November 2020 with optimum constant weighter $c = 1$ is \$850,96 million.

Keywords: Exports, MAPE, non-oil and gas, RMSE, WFTS Lee.

1. PENDAHULUAN

Kegiatan ekspor merupakan sebuah aktivitas perdagangan atau penjualan barang dari dalam negeri ke luar negeri. Perekonomian suatu negara tidak terlepas dari peran penting kegiatan ekspor (Nurawaddah, S. 2019). Ekspor merupakan salah satu mesin pendorong pertumbuhan ekonomi terutama bagi negara berkembang. Pertumbuhan

ekonomi suatu wilayah, dapat dilihat berdasarkan nilai Produk Domestik Regional Bruto (PDRB) yang terdiri dari komponen ekspor dan impor dari sektor migas maupun nonmigas (Ginting, A. M. 2017). Berdasarkan data BPS Prov. Kaltim, sepanjang tahun 2018 sampai 2019 nilai ekspor sektor nonmigas di wilayah Provinsi Kalimantan Timur mengalami penurunan sebesar 917,32 Juta

US\$. Terkait hal ini, maka diperlukan adanya suatu peramalan untuk memperkirakan nilai ekspor dari sektor nonmigas di masa mendatang yang mana sektor tersebut sebagai penunjang nilai PDRB Provinsi Kalimantan Timur.

Peramalan dapat dilakukan dengan menggunakan metode statistika yang salah satunya yaitu *fuzzy time series*. Pada tahun 1993 Song dan Chissom memperkenalkan *fuzzy time series* dengan berdasarkan teori himpunan *fuzzy*, logika *fuzzy*, serta hasil peramalan yang dapat dibahasakan (linguistik). *Fuzzy Time Series* (FTS) senantiasa mengalami perkembangan dari waktu ke waktu. Pada tahun 2005 Hui-Kuang Yu memperkenalkan FTS dengan menggunakan panjang interval sebagai faktor dalam menghasilkan nilai peramalan yang lebih akurat dibandingkan FTS yang telah diperkenalkan oleh Song dan Chissom serta Chen. Di tahun yang sama (2005), Yu memperkenalkan kembali sebuah metode *Weighted Fuzzy Time Series* (WFTS) yang didalamnya terdapat pengulangan dan pembobotan yang berbeda pada tiap pola relasi yang terbentuk.

Pada tahun 2006, Cheng mengembangkan kembali WFTS dengan menggunakan data yang berpola *trend* sedangkan Lee tahun 2009 (Lee, M. H., Efendi, R., & Ismail, Z. 2009). memodifikasi WFTS menggunakan bobot yang berbeda dan berdasarkan nilai tengah dari data aktual. Lee dan Suhartono pada tahun 2010 (Lee, M. H., & Suhartono. 2010). kembali mengembangkan WFTS dengan memperhatikan urutan dari relasi yang terbentuk serta pembobotan yang baru. Hasil peramalan dengan menggunakan metode yang dikembangkan oleh Lee dan Suhartono memiliki akurasi yang lebih baik dibandingkan metode Yu dan Cheng.

2. METODE PENELITIAN

2.1 Peramalan

Peramalan adalah proses memperkirakan atau menduga apa yang akan terjadi di masa mendatang berdasarkan data masa lalu maupun masa kini. Tujuan dari peramalan yaitu sebagai tindakan preventif untuk meminimalisir kejadian yang akan terjadi di masa mendatang.

Berdasarkan jangka waktunya peramalan memiliki 3 jenis yaitu:

1. Peramalan Jangka Pendek
2. Peramalan Jangka Menengah
3. Peramalan Jangka Panjang

Fuzzy time series sendiri termasuk ke dalam peramalan jangka pendek karena hanya dapat

meramalkan sebanyak 1 periode ke depan. Begitu halnya dengan *Weighted Fuzzy Time Series*.

2.2 Logika Fuzzy

Logika *fuzzy* diperkenalkan oleh Prof. Lotfi Astor Zadeh pada tahun 1962. Logika *fuzzy* adalah metodologi sistem kontrol pada pemecahan masalah yang cocok untuk diimplementasikan pada sistem yang sederhana, sistem kecil, dan sistem kontrol. Secara garis besar, logika *fuzzy* merupakan suatu cara untuk mengubah suatu pernyataan dalam bentuk linguistik menjadi nilai numerik. Logika *fuzzy* memungkinkan nilai keanggotaan berada di antara 0 dan 1.

Beberapa hal yang perlu diperhatikan dalam memahami logika *fuzzy* (Jatipaningrum, M. T. 2016) yaitu:

1. Variabel *Fuzzy*.
2. Himpunan *Fuzzy*.
3. Semesta Pembicaraan.
4. *Domain* Himpunan *Fuzzy*.

2.3 Himpunan Fuzzy

Menurut (Nasir, J., & Suprianto, J. (2017). himpunan *fuzzy* merupakan himpunan yang mendefinisikan bahwa suatu pengamatan dapat menjadi bagian dari beberapa himpunan, dengan masing-masing bagian memiliki nilai keanggotaan yang berbeda. Suatu elemen pada himpunan *fuzzy* terdapat lebih dari sama dengan 2 dengan nilai keanggotaannya berada dalam rentang 0 sampai dengan 1. Himpunan *fuzzy* memiliki dua atribut yaitu:

1. Linguistik, yaitu penamaan suatu kelompok yang dapat mewakili suatu keadaan dengan menggunakan bahasa sehari-hari.
2. Numerik, yaitu suatu nilai atau angka yang menyatakan ukuran dari suatu variabel.

2.4 Fuzzy Time Series

Metode *Fuzzy Time Series* (FTS) merupakan salah satu metode statistika yang dapat digunakan untuk melakukan peramalan berdasarkan data masa lalu yang didalamnya termuat logika *fuzzy*, himpunan *fuzzy*, dan hasil peramalan yang dapat dibahasakan (linguistik). Proses peramalan FTS berbentuk nilai linguistik yang berasal dari bilangan *real* terhadap himpunan pembicaraan pada data aktual. Hal inilah yang digunakan sebagai pengganti data masa lalu yang akan diramalkan.

2.5 Weighted Fuzzy Time Series

Weighted Fuzzy Time Series (WFTS) merupakan perkembangan dari FTS yang didalamnya terdapat

$$\hat{F}(t) = \frac{m_{j-1} + m_{j-2} + \dots + m_{j-p}}{p} \quad (11)$$

dengan p adalah banyaknya FLR yang ada dan untuk mencari nilai tengah (m_{j-l}) untuk masing-masing FLR mengikuti Persamaan (5). Sehingga matriks *defuzzifikasi* untuk FLR lebih dari satu dapat dituliskan sebagai berikut:

$$\mathbf{M}(t) = [m_{j-1}, m_{j-2}, \dots, m_{j-p}] \quad (12)$$

- Menghitung nilai pembobot berdasarkan FLRG yang terbentuk. Jika FLRG berasal dari FLR $A_i \rightarrow \emptyset$ dan $A_i \rightarrow A_j$ maka diberi pembobot sebesar 1, sedangkan jika FLRG yang terbentuk adalah $A_{j-1}, A_{j-2}, \dots, A_{j-p}$ maka pembobotnya adalah $w^*(A_{j-1}), w^*(A_{j-2}), \dots, w^*(A_{j-p})$ yang mana

$$w^*(A_{j-l}) = \frac{w(A_{j-l})}{\sum_{l=1}^p w(A_{j-l})} \quad (13)$$

Maka, secara umum $w(A_{j-l})$ dapat dituliskan sebagai berikut :

$$w(A_{j-l}) = \begin{cases} 1, & \text{jika } l = 1 \\ c^{l-1}, & \text{jika } l > 1 \text{ dan } c \geq 1 \end{cases} \quad (14)$$

Berdasarkan persamaan (14) matriks pembobot dapat ditulis :

$$\mathbf{W}_o(t) = \left[\frac{1}{\sum_{l=1}^p w(A_{j-l})}, \frac{c}{\sum_{l=1}^p w(A_{j-l})}, \frac{c^2}{\sum_{l=1}^p w(A_{j-l})}, \dots, \frac{c^{l-1}}{\sum_{l=1}^p w(A_{j-l})} \right] \quad (15)$$

- Peramalan akhir $\hat{X}(t)$ didapatkan dengan perkalian matriks *defuzzifikasi* dengan matriks pembobot. Berikut rumus peramalan akhir:

$$\hat{X}(t) = \mathbf{M}(t) \times \mathbf{W}_o(t)^T \quad (16)$$

2.6 Penentuan Panjang Interval Berbasis Rata-Rata (Average Based)

Berikut langkah penentuan panjang interval berbasis rata-rata :

- Menentukan panjang interval himpunan semesta pembicaraan (U)

$$R = (X_{maks} + D_2) - (X_{min} - D_1) \quad (17)$$

- Menghitung rata-rata selisih absolut setiap data

$$S = \frac{\sum_{t=1}^{N-1} |(X_{t+1}) - X_t|}{N-1} \quad (18)$$

- Menghitung basis interval himpunan *fuzzy*

$$B = \frac{S}{2} \quad (19)$$

Dari hasil perhitungan diatas, maka disesuaikan pada tabel basis interval di bawah untuk menentukan nilai jangkauan yang akan digunakan.

Tabel 1. Basis interval

Jangkauan	Basis
0,1-1	0,1
1,1-10	1
11-100	10
101-1000	100
1001-10000	1000

- Menghitung banyaknya himpunan *fuzzy*
Basis interval digunakan untuk menghitung banyaknya himpunan *fuzzy*. Berikut perhitungan banyaknya himpunan *fuzzy* :

$$n = \frac{R}{B} \quad (20)$$

- Pembentukan himpunan *fuzzy* berdasarkan panjang interval berbasis rata-rata.

2.7 Nilai Keakuratan Peramalan

Pada penelitian ini, untuk mencari nilai keakuratan peramalan digunakan metode *Mean Absolute Percentage Error* (MAPE) dan *Root Mean Square Error* (RMSE). Peramalan dikatakan sangat akurat jika memiliki nilai MAPE di bawah 10% dan dapat dikatakan akurat jika nilainya berada di antara 10% dan 20% sedangkan RMSE dapat diartikan sebagai ukuran kesalahan yang didasarkan pada selisih antara dua buah nilai yang bersesuaian.

Rumus MAPE adalah :

$$\text{MAPE} = \left(\frac{1}{N} \sum_{t=1}^N \left| \frac{X_t - \hat{X}(t)}{X_t} \right| \right) \times 100\% \quad (21)$$

Rumus RMSE adalah :

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^N (X_t - \hat{X}_t)^2}{N}} \quad (22)$$

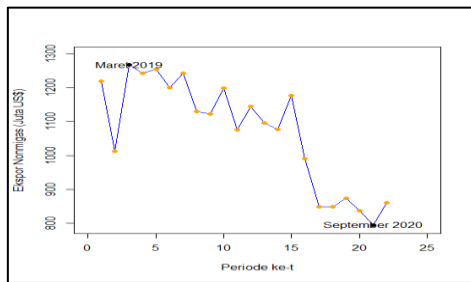
2.8 Ekspor Nonmigas

Ekspor adalah kegiatan perdagangan dengan cara mengirimkan barang dari dalam negeri ke luar negeri dengan memenuhi ketentuan yang berlaku. Melalui ekspor, suatu negara atau daerah dapat menghasilkan devisa yang dapat digunakan sebagai salah satu sumber dana untuk menunjang pembangunan (Fajar, I. S. 2013). Di Indonesia, ekspor terdiri dari beberapa sektor salah satunya sektor migas dan nonmigas. Ekspor migas meliputi barang yang dihasilkan dari minyak dan gas alam sedangkan ekspor nonmigas adalah barang yang

dihasilkan selain minyak dan gas alam. Komoditas ekspor nonmigas terbagi dalam 3 sektor yaitu sektor pertanian, pertambangan, dan industri pengolahan.

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data bulanan ekspor nonmigas Provinsi Kalimantan Timur dari bulan Januari 2019 – Oktober 2020 (Badan Pusat Statistik Provinsi Kalimantan Timur, 2019). Berdasarkan data yang ada, dapat digambarkan *time series plot* seperti pada Gambar 1.



Gambar 1. *Time series plot* ekspor nonmigas

Berdasarkan Gambar 1 dapat dilihat bahwa nilai ekspor nonmigas di Provinsi Kalimantan Timur pada periode Januari 2019 – Oktober 2020 menunjukkan adanya fluktuasi setiap bulannya. Pada bulan Februari 2019, nilai ekspor nonmigas mengalami penurunan yang sangat tinggi dibandingkan penurunan pada bulan-bulan lain sebesar 207,2. Namun, pada bulan berikutnya yaitu bulan Maret 2019 terjadi peningkatan yang cukup drastis dibandingkan peningkatan pada bulan lainnya yaitu sebesar 256,63. Kemudian pada bulan September 2020, nilai ekspor nonmigas mencapai titik terendahnya yaitu sebesar 793,36.

3.1 Penentuan Himpunan Semesta Pembicaraan

Berikut himpunan semesta pembicaraan sesuai Persamaan (1) dengan D_1 dan D_2 merupakan bilangan positif sembarang. Pada penelitian ini digunakan nilai $D_1 = 2,4$ dan $D_2 = 2,39$.

$$\begin{aligned}
 U &= [X_{\min} - D_1; X_{\max} + D_2] \\
 &= [793,36 - 2,4; 1.268,57 + 2,39] \\
 &= [790,96; 1.270,96]
 \end{aligned}$$

3.2 Penentuan Panjang Interval Berbasis Rata-Rata (Average Based)

Adapun perhitungan panjang interval dengan langkah sebagai berikut :

1. Menghitung panjang interval (R) dari himpunan semesta pembicaraan (U)

Panjang interval (R) ditentukan menggunakan Persamaan (17). Berikut perhitungan panjang interval (R) dari semesta pembicaraan :

$$\begin{aligned}
 R &= (X_{\max} + D_2) - (X_{\min} - D_1) \\
 &= (1.268,57 + 2,39) - (793,36 - 2,4) \\
 &= 480
 \end{aligned}$$

2. Menghitung rata-rata selisih absolut setiap data (S)

Selisih absolut data aktual dapat dilihat pada Tabel 2.

Tabel 2. Selisih absolut data aktual

Tahun	Bulan	Ekspor Nonmigas	$ X_t - X_{t-1} $
2019	Januari	1.219,14	-
	Februari	1.011,94	207,2
2020	September	793,36	42,98
	Oktober	860,61	67,25
	Jumlah		1.647,99

Berdasarkan Tabel 2 maka diperoleh jumlah selisih absolut data aktual sebesar 1.647,99. Untuk selanjutnya, dihitung nilai rata-rata selisih absolut setiap data dengan menggunakan Persamaan (18):

$$\begin{aligned}
 S &= \frac{\sum_{t=1}^{N-1} |X_t - X_{t-1}|}{22 - 1} \\
 &= \frac{1.647,99}{22 - 1} \\
 &= 78,48
 \end{aligned}$$

3. Menghitung basis interval himpunan *fuzzy* (B)

Hasil rata-rata selisih absolut setiap data digunakan untuk menghitung basis interval *fuzzy* dengan menggunakan Persamaan (19). Berikut perhitungan basis interval himpunan *fuzzy* :

$$\begin{aligned}
 B &= \frac{S}{2} \\
 &= \frac{78,48}{2} \\
 &= 39,24 \approx 40
 \end{aligned}$$

Berdasarkan hasil perhitungan, diperoleh nilai basis interval yaitu 39,24. Nilai basis interval tersebut, berdasarkan Tabel 1 termasuk dalam jangkauan 11-

100 dengan basis 10. Dengan demikian, nilai interval dapat dibulatkan menjadi 40 sehingga nilai B yang digunakan selanjutnya sebesar 40.

4. Menghitung banyaknya himpunan fuzzy (n)

Setelah diperoleh nilai basis interval sebesar 40 maka selanjutnya akan ditentukan banyaknya himpunan fuzzy berdasarkan Persamaan (20). Berikut perhitungan penentuan banyaknya himpunan fuzzy :

$$n = \frac{R}{B} = \frac{480}{40} = 12$$

5. Setelah diperoleh hasil pada langkah sebelumnya, elanjutnya dapat dituliskan 12 himpunan fuzzy yang memiliki panjang interval yang sama yaitu 40. Himpunan semesta pembicaraan (U) dibagi menjadi 12 himpunan yang sama panjang yaitu u_i dengan $i = 1, 2, \dots, 12$.

Berikut himpunan fuzzy yang terbentuk:

$$\begin{aligned} u_1 &= [790, 96; 830, 96) & u_2 &= [830, 96; 870, 96) \\ u_3 &= [870, 96; 910, 96) & u_4 &= [910, 96; 950, 96) \\ u_5 &= [950, 96; 990, 96) & u_6 &= [990, 96; 1.030, 96) \\ u_7 &= [1.030, 96; 1.070, 96) & u_8 &= [1.070, 96; 1.110, 96) \\ u_9 &= [1.110, 96; 1.150, 96) & u_{10} &= [1.150, 96; 1.190, 96) \\ u_{11} &= [1.190, 96; 1.230, 96) & u_{12} &= [1.230, 96; 1.270, 96) \end{aligned}$$

Setelah diperoleh himpunan fuzzy, dapat ditentukan nilai tengah untuk masing-masing himpunan fuzzy. Perhitungan nilai tengah himpunan fuzzy dilakukan dengan menggunakan Persamaan (5). Berikut hasil perhitungan nilai tengah himpunan fuzzy (m_i) :

Tabel 3. Nilai tengah himpunan fuzzy

Fuzzy Set	m_i
1	810,96
2	850,96
⋮	⋮
11	1.210,96
12	1.250,96

3.3 Pendefinisian Derajat Keanggotaan Himpunan Fuzzy terhadap A_i

Pendefinisian derajat keanggotaan himpunan fuzzy terhadap A_i didasarkan pada 12 himpunan fuzzy yang terbentuk. Nilai fuzzyfikasi dari variabel linguistik data ekspor nonmigas Provinsi

Kalimantan Timur yaitu $A_1, A_2, A_3, \dots, A_{12}$. Setiap himpunan fuzzy u_i dengan $i = 1, 2, 3, \dots, 12$ didefinisikan terhadap A_i dengan menggunakan Persamaan (2).

Setelah diperoleh nilai derajat keanggotaan himpunan fuzzy terhadap A_i , maka selanjutnya dapat diperoleh hasil fuzzyfikasi. Hasil fuzzyfikasi dapat dilihat pada Tabel 4.

Tabel 4. Nilai linguistik hasil fuzzyfikasi

Fuzzyfikasi	Nilai Linguistik
A_1	Sangat sangat turun sekali
A_2	Sangat turun sekali
A_3	Turun sekali
A_4	Cukup turun
A_5	Turun
A_6	Sedikit turun
A_7	Sedikit naik
A_8	Naik
A_9	Cukup naik
A_{10}	Naik sekali
A_{11}	Sangat naik sekali
A_{12}	Sangat Sangat naik sekali

3.4 Fuzzyfikasi Data Aktual

Misal, bulan Januari 2019 dengan nilai ekspor nonmigas sebesar 1.219,14. Nilai tersebut termasuk ke dalam himpunan fuzzy ke-11 (u_{11}) dengan interval $[1.190, 96; 1.230, 96)$, sehingga bulan Januari 2019 memiliki fuzzyfikasi A_{11} .

Tabel 5. Fuzzyfikasi data ekspor nonmigas

Tahun	Bulan	Ekspor Nonmigas	Fuzzyfikasi
2019	Januari	1.219,14	A_6
	Februari	1.011,94	A_{12}
2020	Sept	793,36	A_1
	Oktober	860,61	A_2

3.5 Penentuan Fuzzy Logical Relationship (FLR)

Penentuan FLR berdasarkan current state dan next state dimana current state merupakan data pada pengamatan saat ini dan next state adalah data pengamatan selanjutnya. Misal bulan Januari 2019

merupakan *current state* ($X_{(t-1)}$) dengan nilai *fuzzyfikasi* A_{11} , dan bulan Februari 2019 merupakan *next state* (X_t) dengan nilai *fuzzyfikasi* A_6 , maka FLR yang terbentuk antara bulan Januari 2019 dengan bulan Februari 2019 adalah $A_{11} \rightarrow A_6$.

Tabel 6. FLR

Bulan	FLR
Januari 2019 → Februari 2019	$A_{11} \rightarrow A_6$
Februari 2019 → Maret 2019	$A_6 \rightarrow A_{12}$
⋮	⋮
Agustus 2020 → September 2020	$A_2 \rightarrow A_1$
September 2020 → Oktober 2020	$A_1 \rightarrow A_2$

3.6 Penentuan Fuzzy Logical Relationship Group (FLRG)

FLR yang terbentuk pada Tabel 6 menunjukkan adanya *current state* yang sama meskipun *next state* yang berbeda, sehingga dapat dikelompokkan dalam satu grup yang disebut FLRG. Misal, FLR pada bulan Maret 2019 → April 2019 adalah $A_{12} \rightarrow A_{12}$, begitu pula bulan April 2019 → Mei 2019 adalah $A_{12} \rightarrow A_{12}$, dan bulan Mei 2019 → Juni 2019 yaitu $A_{12} \rightarrow A_{11}$, serta bulan Juli 2019 → Agustus 2019 memiliki FLR $A_{12} \rightarrow A_9$, maka FLRG yang terbentuk yaitu $A_{12} \rightarrow A_{12}, A_{12}, A_{11}, A_9$. FLR yang hanya memiliki satu *next state* ataupun himpunan kosong dapat dikatakan pula sebagai FLRG.

Tabel 7. FLRG

Grup	FLRG
1	$A_1 \rightarrow A_2$
2	$A_2 \rightarrow A_2, A_3, A_1$
⋮	⋮
12	$A_{12} \rightarrow A_{12}, A_{12}, A_{11}, A_9$

3.7 Peramalan Awal dan Defuzzifikasi

Peramalan awal $F(t)$ dan *defuzzifikasi* $\hat{F}(t)$ dilakukan dengan berdasarkan aturan Chen (1996). Peramalan awal adalah proses peramalan dari FLRG yang terbentuk dimana hasil peramalan yang diperoleh masih berbentuk variabel linguistik. Setelah ditentukan peramalan awal, langkah selanjutnya yaitu proses *defuzzifikasi*. Proses *defuzzifikasi* yaitu proses yang bertujuan untuk

mengubah nilai peramalan yang masih berbentuk linguistik menjadi nilai numerik kembali.

Tabel 8. Peramalan awal dan defuzzifikasi

Gru	FLRG	$F(t)$	$\hat{F}(t)$
p			
1	$A_1 \rightarrow A_2$	A_2	850,96
2	$A_2 \rightarrow A_2, A_3, A_1$	A_2, A_3, A_1	850,96
⋮	⋮	⋮	⋮
12	$A_{12} \rightarrow A_{12}, A_{12}, A_{11}, A_9$	$A_{12}, A_{12}, A_{11}, A_9$	1.210,96

3.8 Pembentukan Matriks Defuzzifikasi

Pembentukan matriks *defuzzifikasi* ($M(t)$) didasarkan pada hasil peramalan awal dan *defuzzifikasi* yang telah diperoleh sebelumnya. Pembentukan matriks *defuzzifikasi* mengikuti aturan Chen (1996), dimana matriks *defuzzifikasi* merupakan matriks yang berisikan nilai tengah dari hasil peramalan awal.

Tabel 9. Matriks defuzzifikasi

Gru	$F(t)$	$M(t)$
p		
1	A_2	850,96
2	A_2, A_3, A_1	[850,96 890,96 810,96]
⋮	⋮	⋮
12	$A_{12}, A_{12}, A_{11}, A_9$	[1.250,96 1.250,96 1.210,96 1.130,96]

3.9 Perhitungan Nilai Pembobot

Perhitungan nilai pembobot dilakukan dengan berdasarkan FLRG yang terbentuk. Jika FLRG berasal dari FLR $A_i \rightarrow \emptyset$ atau $A_i \rightarrow A_j$ maka diberi pembobot ($W_o(t)$) sebesar 1. Namun, jika FLRG yang terbentuk adalah $A_i \rightarrow A_{j-1}, A_{j-2}, \dots, A_{j-p}$ maka pembobot untuk masing-masing *next state* adalah $w^*(A_{j-1}), w^*(A_{j-2}), \dots, w^*(A_{j-p})$ sebagaimana pada Persamaan (13) dengan nilai konstanta pembobot (c) yang digunakan yaitu $1 \leq c \leq 2$ dan jarak antar bobot sebesar 0,1. Sebagai contoh pada FLRG dari grup ke-2 yaitu $A_2 \rightarrow A_2, A_3, A_1$ maka dapat ditentukan pembobotnya ialah $w^*(A_2), w^*(A_3), w^*(A_1)$. Sebelum dilakukan perhitungan pembobot terlebih dahulu dihitung nilai

$w(A_{j-1})$ untuk setiap *next state* FLRG grup ke-2 berdasarkan Persamaan (14) :

$$w(A_{j-1}) = 1$$

$$w(A_{j-2}) = c^{l-1} = c^{2-1} = c$$

$$w(A_{j-3}) = c^{l-1} = c^{3-1} = c^2$$

dengan demikian, $\sum_{l=1}^3 w(A_{j-l}) = 1 + c + c^2$ sehingga dapat diperoleh pembobot masing-masing *next state* berdasarkan Persamaan (13) berikut :

$$w^*(A_2) = \frac{1}{\sum_{l=1}^3 w(A_{j-l})} = \frac{1}{1+c+c^2}$$

$$w^*(A_3) = \frac{c^{l-1}}{\sum_{l=1}^3 w(A_{j-l})} = \frac{c^{2-1}}{\sum_{l=1}^3 w(A_{j-l})} = \frac{c}{\sum_{l=1}^3 w(A_{j-l})} = \frac{c}{1+c+c^2}$$

$$w^*(A_1) = \frac{c^{l-1}}{\sum_{l=1}^3 w(A_{j-l})} = \frac{c^{3-1}}{\sum_{l=1}^3 w(A_{j-l})} = \frac{c^2}{\sum_{l=1}^3 w(A_{j-l})} = \frac{c^2}{1+c+c^2}$$

atau dapat dituliskan dalam matriks pembobot sebagaimana pada Persamaan (15).

$$W_o(2) = \begin{bmatrix} 1 & c & c^2 \\ \sum_{l=1}^p w(A_{j-l}) & \sum_{l=1}^p w(A_{j-l}) & \sum_{l=1}^p w(A_{j-l}) \end{bmatrix}$$

Jika $c=1$ maka dapat diperoleh nilai matriks pembobot :

$$W_o(2) = \begin{bmatrix} 1 & 1 & 1^2 \\ 3 & 3 & 3 \end{bmatrix}$$

Perhitungan nilai pembobot untuk setiap kontanta pembobot (c) yang berbeda yaitu $1 \leq c \leq 2$ dan grup lainnya mengikuti seperti pada perhitungan $c=1$ pada grup ke-2.

Tabel 10. Nilai pembobot

Grup	$c=1$...	$c=2$
1	1	...	1
2	$\frac{1}{3}; \frac{1}{3}; \frac{1}{3}$...	$\frac{1}{7}; \frac{2}{7}; \frac{4}{7}$
⋮	⋮	⋮	⋮
12	$\frac{1}{4}; \frac{1}{4}; \frac{1}{4}; \frac{1}{4}$...	$\frac{1}{15}; \frac{2}{15}; \frac{4}{15}; \frac{8}{15}$

3.10 Peramalan Akhir

Setelah mendapatkan nilai pembobot berdasarkan nilai konstanta pembobot (c) yang

berbeda-beda, dan juga matriks defuzzifikasi ($M(t)$), maka selanjutnya dapat dilakukan peramalan akhir yang merupakan proses akhir dari WFTS Lee (2010).

Tabel 11. Peramalan akhir

Gr up	$\hat{X}(t)$			Periode Peramalan
	$c=1$...	$c=2$	
1	850,96	...	850,96	Oktober 2020
2	850,96	...	838,65	Juni 2020, Juli 2020, dan September 2020
⋮	⋮	⋮	⋮	⋮
12	1.210,96	...	1.173,72	April 2019, Mei 2019, Juni 2019, dan Agustus 2019

Selanjutnya, untuk memudahkan pembacaan hasil peramalan akhir maka dilakukan penyusunan berdasarkan bulan.

Tabel 12. Peramalan akhir Januari 2019 – Oktober 2020

Tahun	Bulan	X_t	$\hat{X}(t)$		
			$c=1$...	$c=2$
2019	Jan	1.219,14	-	...	-
	Feb	1.011,94	1.117,63	...	1.134,04
⋮	⋮	⋮	⋮	⋮	⋮
2020	Sept	793,36	850,96	...	838,65
	Okt	860,61	850,96	...	850,96

Berdasarkan Tabel 12 selanjutnya dapat dihitung nilai ketepatan akurasi peramalan dengan menggunakan metode MAPE sebagaimana pada Persamaan (21) dan metode RMSE pada Persamaan (22).

Tabel 13. Nilai akurasi peramalan

Nilai c	RMSE	MAPE
$c=1$	50,67	3,62%
$c=1,1$	50,74	3,62%
⋮	⋮	⋮
$c=2$	53,37	3,82%

Tabel 13 menyatakan bahwa nilai konstanta pembobot $c = 1$ dan $c = 1,1$ memiliki nilai akurasi MAPE yang sama yaitu sebesar 3,62%, nilai ini lebih kecil dibanding nilai c lainnya. Namun, dengan metode akurasi peramalan yang berbeda yaitu metode RMSE nilai konstanta pembobot $c = 1$ memiliki nilai yang minimum dibanding nilai konstanta pembobot lainnya yaitu sebesar 50,67. Secara keseluruhan, dapat dinyatakan bahwa akurasi peramalan dari WFTS Lee untuk data ekspor Nonmigas Provinsi Kalimantan Timur dari bulan Januari 2019 – Oktober 2020 memiliki keakuratan yang sangat baik dengan kontanta pembobot terbaiknya yaitu $c = 1$.

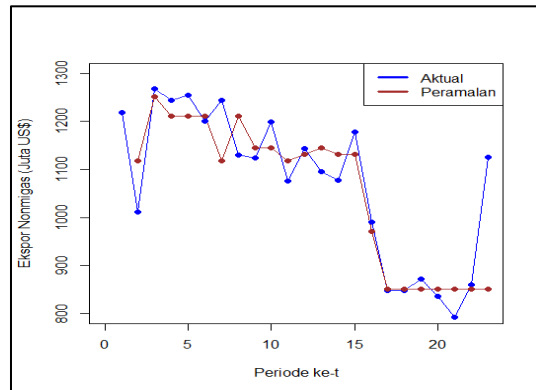
Berdasarkan hasil nilai konstanta pembobot terbaik yang telah diperoleh, maka selanjutnya dapat dilakukan peramalan untuk satu bulan ke depan yaitu bulan November 2020. Pada peramalan bulan November 2020 (sebagai *next state*) maka sebagai *current state* yaitu bulan Oktober 2020 yang terletak pada *fuzzyfikasi* A_2 (Tabel 5). Nilai *fuzzyfikasi* dari A_2 membentuk FLRG $A_2 \rightarrow A_2, A_3, A_1$ dan termasuk ke dalam *defuzzifikasi* grup FLRG ke-2. Hasil peramalan akhir untuk grup ke-2 sebagaimana pada Tabel 11 yaitu sebesar 850,96. Sehingga, hasil peramalan bulan November 2020 adalah 850,96. Peramalan bulan November 2020 dengan nilai konstanta pembobot (c) lainnya mengikuti langkah yang sama seperti pada $c=1$ dan secara lengkap dapat dilihat pada Tabel 14

Tabel 14. Hasil peramalan akhir bulan November 2020

Nilai c	November 2020
$c=1$	850,96
$c=1,1$	849,59
$c=1,2$	848,2
$c=1,3$	846,81
$c=1,4$	845,46
$c=1,5$	844,17
$c=1,6$	842,93
$c=1,7$	841,77
$c=1,8$	840,67
$c=1,9$	839,63
$c=2$	838,65

Berikut disajikan *time series plot* perbandingan hasil peramalan akhir dan data aktual ekspor Nonmigas Provinsi Kalimantan Timur bulan Januari 2019 – November 2020 dengan menggunakan metode WFTS Lee dengan nilai konstanta

pembobot (c) terbaik yaitu $c = 1$ dapat dilihat pada Gambar 3.



Gambar 3. *Time series plot* perbandingan hasil peramalan dengan data aktual

KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan, maka kesimpulan yang dapat diperoleh dari penelitian ini yaitu :

1. Hasil peramalan data ekspor nonmigas Provinsi Kalimantan Timur pada bulan November 2020 dengan menggunakan metode *Weighted Fuzzy Time Series* Lee (2010) dengan kontanta pembobot terbaik $c = 1$ adalah sebesar 850,96.
2. Berdasarkan pembahasan dengan metode WFTS Lee diperoleh nilai akurasi peramalan dengan menggunakan metode MAPE dan metode RMSE dengan konstanta pembobot (c) yaitu $1 \leq c \leq 2$ dan diperoleh nilai MAPE keseluruhan berada dibawah 10% dengan nilai minimum pada $c = 1$ sebesar 3,62% sama halnya dengan nilai RMSE memiliki nilai minimum pada $c = 1$ yaitu sebesar 50,67.

DAFTAR PUSTAKA

Badan Pusat Statistik Provinsi Kalimantan Timur. (2019). *Statistik Ekspor Provinsi Kalimantan Timur*. Badan Pusat Statistik Provinsi Kalimantan Timur.

Fajar, I. S. (2013). Pengaruh Ekspor-Impor dan Indeks Harga Konsumen (IHK) Terhadap Pertumbuhan Ekonomi Indonesia. *Jurnal Ilmiah Mahasiswa Ekonomi, Bisnis & Manajemen*.

Ginting, A. M. (2017). Analisis Pengaruh Ekspor Terhadap Pertumbuhan Ekonomi Indonesia. *Buletin Ilmiah Litbang Perdagangan*, 11(1), 1-

20.

Jatipaningrum, M. T. (2016). Peramalan Data Produk Domestik Bruto dengan Fuzzy Time Series Markov Chain. *Jurnal Teknologi*, 9(1), 31–38.

Lee, M. H., Efendi, R., & Ismail, Z. (2009). Modified Weighted for Enrollment Forecasting Based on Fuzzy Time Series. *Matematika*, 25(1), 67-69

Lee, M. H., & Suhartono. (2010). A Novel Weighted Fuzzy Time Series Model For Forecasting Seasonal Data. *International Conference on Mathematical Sciences*, 332–340.

Nasir, J., & Suprianto, J. (2017). Analisis Fuzzy Logic Menentukan Pemilihan Motor Honda dengan Metode Mamdani. *Jurnal Edik Informatika*, 3(2), 177–187.

Nurmawaddah, S. (2019). Pengaruh Ekspor Non Migas terhadap Pertumbuhan Ekonomi Provinsi Kalimantan Timur. *Akuntabel*.16(1), 128-133.

Peramalan Indeks Harga Saham PT Verena Multi Finance Tbk Dengan Metode Pemodelan ARIMA Dan ARCH-GARCH

Fajriyah Fitriyani⁽¹⁾, Salsabila Fasya A⁽²⁾, Muhammad Irfan R⁽³⁾, Teguh Ammar T⁽⁴⁾,
Universitas Padjadjaran
Gedung Departemen Statistika, Jatinangor 45361, Sumedang, Jawa Barat
e-mail: muhammad18011@mail.unpad.ac.id

ABSTRAK

Indeks harga saham merupakan indikator penting dalam mencerminkan keseluruhan pergerakan atau naik turunnya harga saham yang ada dalam suatu periode. Indeks harga saham tentunya memiliki fungsi sebagai ukuran trend pasar, dimana menggambarkan keadaan pasar pada saat tertentu. Acuan kinerja suatu pasar modal serta produk hasil investasi dapat diketahui dengan ukuran indeks saham. Oleh karena itu, penelitian ini akan memfokuskan peramalan pada indeks harga saham PT Verena Multi Finance Tbk yang merupakan perusahaan yang bergerak di bidang pembiayaan konsumen. Masih kurangnya keakuratan metode peramalan yang digunakan sebelumnya menjadi latar belakang dari penelitian ini untuk mendapatkan prediksi indeks harga saham PT Verena Multi Finance Tbk dengan metode yang memiliki ketepatan yang tinggi. Data yang dipakai dalam penelitian ini merupakan data sekunder yang didapat dari website *yahoofinance.com* periode penelitian yaitu Januari 2009 sampai dengan Desember 2019. Metode yang digunakan pada penelitian ini merupakan metode ARIMA, yaitu merupakan metode yang dipakai untuk peramalan data *time series*. Berdasarkan pengolahan data, didapat model terbaik dengan nilai AIC terkecil yaitu ARIMA (0,1,1). Metode yang selanjutnya digunakan yaitu model ARCH-GARCH yang merupakan metode dalam peramalan data yang memiliki masalah heteroskedastisitas. Berdasarkan hasil analisis yang diperoleh dengan peramalan ARCH-GARCH, didapatkan nilai indeks harga saham pada bulan Desember 2021 adalah sebesar 102.4 dengan MAPE sebesar 22.9971%. Dengan melakukan penelitian peramalan indeks harga saham ini, diharapkan dapat ditemukan model paling baik yang dapat meramalkan indeks harga saham PT Verena Multi Finance Tbk pada waktu-waktu berikutnya agar dapat menjadi tolak ukur atau sebagai acuan kinerja pasar modal serta produk hasil investasi serta rujukan yang dapat dipercaya untuk melihat kondisi bursa saham pada saat ini.

Kata kunci : *Indeks harga saham, peramalan, ARIMA, ARCH-GARCH.*

ABSTRACT

*The stock price index is an important indicator in reflecting the overall movement of stock prices in a period. This index certainly has a function as an indicator of market trends, which describe market conditions at a certain time. The stock index is very important as a benchmark for the performance of the capital market and investment products. Therefore, this study will focus on forecasting the stock price index of PT Verena Multi Finance Tbk, which is a company engaged in consumer finance. The lack of accuracy of the forecasting method used previously became the background of this study to get a prediction of the monthly period of PT Verena Multi Finance Tbk's stock price index with a method that has higher accuracy. The data used is secondary data obtained from *yahoofinance.com* for the period January 2009 to December 2019. The method used in this study is the ARIMA model, which is one of the methods used for forecasting time series data. Based on data processing, the best model with the smallest AIC value was obtained, namely ARIMA(0,1,1). The next method used is the ARCH-GARCH model which is a method used in forecasting data that has heteroscedasticity problems without eliminating the heteroscedasticity. Based on the analysis results obtained by the ARCH-GARCH forecasting, the stock price index value in December 2021 was 102.4 with a MAPE of 22,9971%. By conducting this research, it is hoped that the best model can be found to predict the stock price index of PT Verena Multi Finance Tbk in the following times so that it can be a benchmark for the performance of the capital market and investment products as well as a reliable reference to see the current condition of the stock market.*

Keywords : *Stock price index, Forecasting, ARIMA, ARCH-GARCH.*

1. PENDAHULUAN

Pasar modal merupakan suatu tempat dimana dua pihak bertemu, yaitu emiten dan investor yang didalamnya terdapat sebuah kegiatan berhubungan dengan perdagangan dan juga penawaran umum. Keberadaan pasar modal memberikan perusahaan agar dapat memanfaatkan dana yang diperoleh untuk dipakai sebagai bahan pengembangan dalam sebuah penambahan modal, dan kegiatan usaha lainnya. Pasar modal juga dapat dijadikan sebagai sarana alternatif bagi masyarakat dalam menginvestasikan harta kekayaannya yang berupa uang dengan harapan yang besar serta mampu memberikan keuntungan dengan risiko yang sekecil mungkin. Peran pasar modal di atas diharapkan dapat memakmurkan masyarakat serta membantu meningkatkan perekonomian suatu negara.

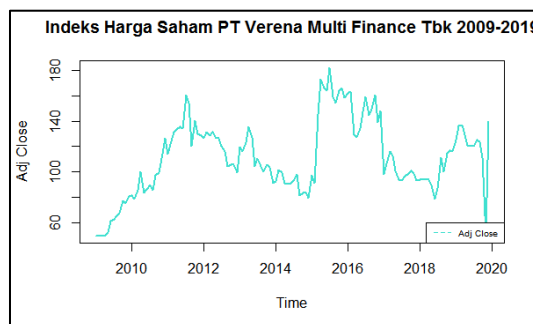
Salah satu investasi yang sering dilakukan di pasar modal adalah saham. Saham adalah suatu surat yang dijadikan sebagai bukti saat seseorang mempunyai atau memegang bagian modal dalam suatu perusahaan. Seseorang dengan kepemilikan saham tentunya mempunyai hak terhadap sebagian aset perusahaan. Salah satu cara perusahaan untuk memperoleh modal atau dana untuk mengembangkan bisnis jangka panjang adalah dengan melakukan penerbitan saham, yang dimana saham sendiri dapat diperjualbelikan di bursa saham dengan harga yang berfluktuasi sesuai keadaan perusahaan serta ekonomi.

Indikator yang mencerminkan keseluruhan Bergeraknya atau naik turunnya harga saham di suatu periode disebut juga Indeks harga saham. Indeks ini berfungsi sebagai ukuran trend pasar, yang melukiskan keadaan pasar pada saat tertentu. Saat indeks naik, maka saham-saham yang masuk ke dalam indeks pun turut naik dan sebaliknya. Gambar yang ditampilkan dari indeks harga saham dapat dijadikan sebagai suatu pedoman atau acuan para investor untuk menghitung serta menentukan tingkat pengembalian waktu ke waktu. Tingkat pengembalian ini dapat ditentukan dengan membandingkan tingkat indeks saat ini dengan tingkat indeks pada masa lampau.

Penelitian ini mengenai peramalan indeks harga saham memperoleh perhatian yang baik untuk para peneliti pada bidang makroekonomi. Peramalan indeks harga saham dapat dijadikan sebagai tolak ukur atau acuan dalam suatu kinerja pasar modal serta produk hasil investasi, karena dengan adanya indeks ini dapat memberikan kemudahan bagi investor dalam memantau perkembangan pasar saham tanpa harus selalu

memerhatikan naik turunnya tiap saham. Penelitian indeks harga saham ini juga dapat dijadikan sebagai suatu rujukan yang dapat dipercaya untuk melihat kondisi bursa saham pada saat ini.

Pada kali ini akan dibahas mengenai indeks harga PT Verena Multi Finance Tbk, yaitu salah satu perusahaan yang bergerak di dalam bidang pembiayaan konsumen periode Januari 2009 – Desember 2019.



Gambar 1. Plot indeks harga saham PT Verena Multi Finance Tbk Januari 2009 – Desember 2019

Terlihat pada Gambar 1. bahwa plot Indeks Harga Saham PT Verena Multi Finance Tbk pada periode Januari 2009 – Desember 2019 menunjukkan data yang tidak stabil atau tingkat variabilitas yang berubah-ubah dari waktu ke waktu. Enders (1995) mengemukakan bahwa data *time series* dengan ragam tidak konstan disebut sebagai data *time series* dengan heteroskedastisitas bersyarat (*conditional heteroskedastic*). Dalam hal ini dibutuhkan suatu metode peramalan yang tepat untuk mengatasi kemungkinan ketidakstabilan data, yaitu tingkat variabilitas yang berubah-ubah dari waktu ke waktu. Metode yang digunakan harus dapat memodelkan sebagian besar data dengan tetap mempertahankan heteroskedastisitas data. Berdasarkan uraian di atas, untuk data yang mengandung heteroskedastisitas seperti pada Indeks Harga Saham PT Verena Multi Finance Tbk, menurut Bollerslev dapat dianalisis menggunakan metode statistik yaitu *Generalized Autoregressive Conditional Heteroscedasticity*.

2. METODE PENELITIAN

Data Penelitian

Di penelitian ini data yang akan dipakai merupakan data sekunder yang didapatkan melalui website *yahoofinance.com*. Data tersebut merupakan data Indeks Harga Saham PT Verena Multi Finance Tbk dengan periode bulanan yang dimulai pada bulan Januari 2009 sampai dengan bulan Desember tahun

2019. Beberapa metode statistik yang dapat digunakan dalam suatu peramalan data deret waktu dimana peramalan dilakukan dengan memakai data pada masa lalu untuk meramalkan kejadian dimasa berikutnya. Penggunaan metode peramalan harus berdasar pada pola datanya. Metode analisis yang digunakan dalam penelitian ini adalah sebagai berikut.

Analisis Runtun Waktu

Runtun waktu merupakan himpunan observasi data yang terurut dalam waktu (Hanke & Wichern, 2005). Metode deret waktu merupakan metode yang menganalisis relasi antara variabel yang akan diperkirakan dengan variabel periode atau waktu. Data runtun waktu atau *time series* merupakan salah satu jenis data yang terdiri dari variabel-variabel yang telah dikumpulkan yang terurut dalam rentang waktu tertentu misalnya mingguan, semesteran dan lain-lain. Pola dalam analisis deret waktu dapat dibagi menjadi empat jenis, yaitu pola data horizontal, trend, musiman, dan siklis (Hanke & Wichern, 2005).

Untuk dapat mengetahui apakah peristiwa yang diamati mengikuti pola-pola perkembangan yang teratur atau tidak, dapat dilihat dari rangkaian waktunya. Apabila pola yang didapat teratur, artinya dapat dibuat peramalan yang bisa dibilang kuat mengenai tingkah laku gejala dari data yang sedang dianalisis, sehingga dapat membuat rencana-rencana di masa yang berikutnya.

ARIMA

ARIMA merupakan model yang dengan penuh mengabaikan variabel independen dalam melakukan suatu peramalan dan suatu model yang berasumsi bahwa data input harus bersifat stasioner (Wei W., 1990). ARIMA memakai data masa lampau dan juga masa sekarang dari variabel dependen untuk melakukan peramalan jangka pendek dengan hasil yang akurat dalam sebuah peramalan. Suatu proses ARIMA(p,d,q) didefinisikan dengan persamaan :

$$D_p(B)(1 - B)^d X_t = C_q(B)\epsilon_t \quad (1)$$

Dengan :

$D(B) = 1 - a_1B - a_2B^2 - a_3B^3 - \dots - a_pB^p$ (operator AR)

$C(B) = 1 - b_1B - b_2B^2 - b_3B^3 - \dots - a_qB^q$ (operator MA)

dengan

X_t = nilai variabel pada waktu ke-t

B = operator backshift

$(1-B)^d X_t$ = deret waktu stasioner pada pembedaan ke-d

ϵ_t = nilai galat pada waktu ke-t

Fajriyah Fitriyani¹, Salsabila Fasya A², Muhammad Irfan R³, Teguh Ammar T⁴/

p = order AR

d = order differencing

q = order MA

Tahap pertama yang dilakukan yaitu uji asumsi meliputi uji stasioneritas dalam rata-rata dan uji stasioneritas dalam varians. Selanjutnya, indentifikasi model dan penaksiran parameter. Kemudian pemilihan model ARIMA terbaik. Setelah didapatkan model ARIMA terbaik, akan dilakukan pengujian efek heteroskedastisitas terhadap model, pendugaan parameter dan pemilihan model ARCH-GARCH terbaik. Selanjutnya model yang didapatkan akan digunakan untuk memprediksi indeks saham.

1. Stasioneritas Data

Menurut Makridakis (1999) agar dapat dilakukan analisis yaitu memakai metode ARIMA, suatu data *time series* wajib berasumsi stasioneritas. Asumsi ini digunakan untuk memperkecil kekeliruan dari model yang dibentuk. Ada 2 jenis bentuk kestasioneran data, yaitu stasioneritas dalam *mean* atau rata-rata dan juga stasioneritas dalam varians. Hal pertama yang harus dilakukan yaitu melakukan plot data indeks saham dari periode penelitian yaitu bulan tahun 2009 sampai bulan Desember 2019.

Dari plot data, dapat dilihat secara visual dan diidentifikasi apakah data bersifat stasioneritas atau tidak. Suatu proses tidak stasioner dalam rata-rata apabila terdapat tren (menurun atau naik) dalam data. Sedangkan ketidakstasioneran dapat dilihat apabila terdapat heteroskedastisitas dalam pergerakan data. Data runtun waktu dapat disebut telah stasioner jika mean data fluktuasi pada sekitar nilai ratarata yang tetap dari waktu ke waktu, sedangkan data runtun waktu dapat disebut telah stasioner dalam varian apabila data fluktuasi dengan ragam yang tetap dari waktu ke waktu (Juanda, Bambang, & Junaidi, 2012).

Namun, stasioneritas berdasarkan plot data akan menghasilkan penilaian yang subjektif. Stasioneritas data indeks saham juga dapat dilaksanakan dengan uji unit root menggunakan metode yaitu *Augmented Dickey-Fuller Test*. Berikut model yang dapat dipakai untuk Uji ADF :

$$\Delta Y_t = \delta Y_{t-1} + u_t \quad (\text{tanpa } intercept)$$

$$\Delta Y_t = \beta + \delta Y_{t-1} + u_t \quad (\text{dengan } intercept)$$

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t \quad (\text{intercept trend waktu})$$

• Hipotesis penelitian

H_0 : $\delta = 0$ (terdapat *unit root*, data tidak memenuhi asumsi stasioner)

H_1 : $\delta \neq 0$ (tidak terdapat *unit root*, data stasioner atau memenuhi asumsi)

• Statistik Uji

$$\hat{\tau} = \frac{\delta}{SE\delta} \quad (2)$$

- Kriteria Uji

Tolak H_0 jika $p\text{-value} < \alpha$, artinya data telah stasioner, terima H_0 dalam hal lainnya.

2. Melakukan Identifikasi Model

Langkah selanjutnya melakukan identifikasi model yang dilakukan pada data yang telah bersifat stasioner. Jika data yang dipakai masih belum memenuhi syarat stasioneritas, maka data tersebut harus dilakukan differencing apabila data tidak stasioner dalam rata-rata atau mean dan dilakukan transformasi apabila belum stasioner dalam varians. Selanjutnya, setelah data yang digunakan bersifat stasioner, dapat ditentukan model yang berbentuk ARMA atau kepanjangan dari *Autoregressive Moving Average* yang akurat dalam penggambaran sifat data, hal ini dilakukan dengan melakukan perbandingan plot pada sampel ACF/PACF dengan sifat fungsi ACF/PACF yang teoritis. Dalam sebuah proses penentuan model yang akurat tentunya dilaksanakan dengan mengidentifikasi orde AR dan MA pada plot ACF dan juga plot PACF. Dalam menentukan orde pada model ARIMA (p,d,q), orde p dan q dapat dilihat dari plot ACF/PACF dan untuk orde d didapatkan dari banyaknya proses differencing yang dilakukan untuk memperoleh data yang telah stasioner dalam rata-rata. Plot ACF berguna untuk mengidentifikasi orde MA dan plot PACF berguna untuk mengidentifikasi orde AR.

Identifikasi orde AR dan juga MA dapat dilaksanakan sesuai dengan correlogram yang diperoleh. Jika ACF turun secara eksponensial dan PACF signifikan pada lag p maka model yang diperoleh adalah AR(p), sebaliknya jika PACF turun secara eksponensial dan ACF signifikan pada lag q maka model yang diperoleh adalah MA(q). Jika ACF turun dengan eksponensial setelah lag q dan PACF turun dengan eksponensial setelah lag p maka bentuk model yang diperoleh adalah ARMA(p,q).

3. Penaksiran Parameter

Apabila telah menentukan bentuk model yang sekiranya sudah sesuai untuk data, maka berikutnya akan dicari nilai penaksir terhadap parameter yang ada di dalam model, contohnya koefisien model ARMA dan nilai variansi residual. Untuk mencari penaksir model ARMA dapat diketahui menggunakan metode Maksimum Likelihood Estimator (MLE), Least Square, Hannan Rissanen, dan lainnya. Selanjutnya, untuk

menguji apakah koefisien hasil taksiran signifikan atau tidak, gunakanlah pengujian dengan menggunakan statistik uji t yang berdistribusi student-t dengan derajat bebas n-1, n = banyaknya sampel. Apabila ada koefisien yang tidak signifikan, maka koefisien tersebut dapat dibuang dari model dan setelah itu estimasi kembali model dengan tidak melibatkan order yang tidak signifikan. Apabila θ merupakan parameter model ARIMA dan $\hat{\theta}$ merupakan nilai estimasi atau taksiran parameter tersebut, dan $S(\hat{\theta})$ merupakan standar error nilai estimasi atau taksiran $\hat{\theta}$, uji signifikansi parameter dapat dilaksanakan seperti langkah berikut.

- Hipotesis penelitian

$H_0 : \theta = 0$ (artinya parameter model ARIMA Box-Jenkins tidak signifikan)

$H_1 : \theta \neq 0$ (artinya parameter model ARIMA Box-Jenkins signifikan)

- Statistik uji

$$\hat{\tau} = \frac{\delta}{SE\delta} \quad (2)$$

- Kriteria uji

Tolak H_0 apabila $|t_{hitung}| \geq t_{tabel}(\alpha, n - p)$ dengan p merupakan banyak parameter, terima H_0 dalam hal lainnya.

4. Pemeriksaan Asumsi

Apabila sudah menaksir nilai-nilai parameter dari model ARIMA, langkah yang harus dilakukan berikutnya adalah *diagnostic check* model yang sudah kita estimasi sebelumnya dengan cara memverifikasi model sesuai sifat pada data. Apabila model sudah tepat, maka data yang dihitung dengan model, pastinya akan memiliki sifat yang tidak beda jauh daripada data aslinya. Residual yang telah dihitung didapatkan dari model yang sudah diestimasi yang memenuhi asumsi dari eror dari model teoritis, seperti normalitas residual, White Noise, dan lain sebagainya. Untuk mengetahui residual berautokorelasi atau tidak, hal tersebut dapat diketahui dengan cara melakukan uji korelasi serial, yaitu menggunakan uji Q-Ljung Box (Box & Pierce, 1970).

A. Uji Q-Ljung Box

- Hipotesis penelitian

$H_0 : \rho_1 = \rho_2 = \rho_3 = \dots = \rho_k = 0$

(artinya residual tidak berautokorelasi)

$H_1 : \text{paling tidak ada satu } \rho_k \neq 0$

(artinya residual berautokorelasi)

- Statistik uji

$$Q = n'(n' + 2) \sum_{k=1}^m \frac{rk^2}{n'} \quad (3)$$

Mengikuti distribusi Chi-Kuadrat dengan $db = (k - p - q - P - Q)$, dan $n' = n - (d + SD)$.

d : orde pembedaan (differencing) bukan faktor *seasonal*

D : orde pembedaan (differencing) faktor *seasonal*

S : lag periode maksimum

rk : autokorelasi time lag 1,2, 3 ... , k .

- Kriteria uji
Tolak H_0 apabila $Q > \chi^2(\alpha, db)$ atau $p - value < \alpha$, terima H_0 dalam hal lainnya.
- B. Uji Kolmogorov-Smirnov
- Hipotesis penelitian
 H_0 : $F(\alpha_t) = F_0(\alpha_t)$ (Residualnya berdistribusi normal)
 H_1 : $F(\alpha_t) \neq F_0(\alpha_t)$ (Residualnya tidak berdistribusi normal)
- Statistik uji
 $D = \max|F_0(x) - S_N(x)|$ (4)
dengan :
 D : difference absolute
 $F_0(x)$: fungsi distribusi kumulatif teoritis dibawah Hipotesis nol
 $S_N(x)$: distribusi frekuensi kumulatif yang diobservasi dari suatu sampel random
- Kriteria uji
Tolak H_0 jika *nilai D > nilai D*(alpha)*, terima dalam hal lainnya. $D^*(\alpha)$ disini merupakan nilai kritis yang didapatkan dari kolmogorov-smirnov tabel.

5. Pemilihan Model ARIMA terbaik

Apabila terdapat 2 metode peramalan yang digunakan pada suatu data, pilihlah model terbaik berdasarkan pada error peramalan (Juanda, Bambang, & Junaidi, 2012). Pemilihan model terbaik adalah metode yang memberikan kesalahan peramalan terkecil. Untuk memilih model terbaik, pilihlah model yang mengecilkkan nilai kriteria Akaike Information Criteria (AIC), yang akan digunakan pada penelitian ini. Berikut persamaannya.

$$AIC : AIC = n \ln(\sigma \varepsilon^2) + 2(m) \tag{5}$$

$$\sigma \varepsilon^2 = \frac{SSE}{n} \tag{6}$$

Dengan,

- n : banyaknya pengamatan
- m : banyaknya parameter dalam model
- SSE : merupakan singkatan dari Sum of Squared Error. SSE bisa ditaksir dari jumlah

Fajriyah Fitriyani¹, Salsabila Fasya A², Muhammad Irfan R³, Teguh Ammar T⁴/

kuadrat seluruh residual. Semakin kecil nilai Akaike Information Criteria (AIC) maka semakin baik model yang didapatkan.

ARCH

Model ARCH atau kepanjangan dari *Autoregressive conditional Heteroscedasticity* merupakan hasil pengembangan Engle. Rata-rata serta varian suatu data dengan deret waktu yang dimodelkan simultan dengan varian residual yang berbeda-beda karena adanya perubahan ragam residual yang tergantung dari seberapa besar residual di masa lampau.

Pada model ARCH(m), diperoleh variansi dari r_t kondisional terhadap informasi masa lalu sebagai.

$$Var(r_t|\zeta_{t-1}) = E(\varepsilon_t^2|F_{t-1}) = \sigma_t^2$$

dapat digambarkan dengan persamaan:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i \varepsilon_{t-i}^2 \tag{7}$$

dengan $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, 2, \dots, m$. Kondisi $\alpha_i \geq 0$ diperlukan agar persamaan vitalitas tidak negatif. Jika semua $\alpha_i = 0$, maka variansi bersyarat σ^2 akan menjadi konstanta α_0 dan ε_t akan merupakan proses IID yang bersifat homoskedastik.

Pada model ARCH, proses ε_t menunjukkan komponen random dari model (sering disebut sebagai proses White Noise), dimana $(\varepsilon_t) = 0$ dan bersifat tidak berkorelasi dengan waktu lampau atau waktu yang akan datang, yakni sebagai berikut.

$$E(\varepsilon_t \varepsilon_s) = \begin{cases} 0, & t \neq s \\ \sigma_t^2, & t = s \end{cases}$$

Dari persamaan di atas walaupun proses ε_t sifatnya tidak saling berkorelasi namun proses tersebut tidak bersifat independen. Dalam model ARCH(m), proses ε_t dibangkitkan oleh proses:

$$\varepsilon_t = \sigma_t v_t$$

dimana σ_t adalah akar positif dari σ^2 dan v_t adalah suatu proses IID (independent dan identically distributed), v_t merupakan white noise dimana rata-rata atau mean-nya sama dengan nol dan variansi satu, dan v_t sering diasumsikan normal standar $N(0,1)$ atau berdistribusi student-t, dan lebih lanjut, v_t independen dengan ε_t , $s < t$. Asumsi ini mengakibatkan distribusi bersyarat dari ε_t akan berdistribusi normal dengan mean 0 dan variansi σ^2 (Wei W., 1990). Model ARCH(p) bisa ditulis dalam bentuk sebagai berikut:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 \tag{8}$$

GARCH

Model GARCH merupakan model hasil pengembangan Bollerslev yang muncul tahun 1986. Model ini adalah model kelanjutan dari

ARCH. Tujuan model GARCH dibangun pada model ARCH adalah untuk menghindari adanya ordo yang tingginya berlebihan. Sesuai dengan prinsip parsimony yaitu dengan memilih model yang lebih sederhana, maka variansinya akan selalu positif. (Model GARCH (p,q) dapat dinyatakan dalam bentuk persamaan berikut :

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 + \lambda_1 \sigma_{t-1}^2 + \dots + \lambda_q \sigma_{t-q}^2 \quad (9)$$

Dimana,

σ_t^2 = kondisional varians

α_0, α_1 = konstanta

ε_{t-p}^2 = residual bulan t-p

σ_{t-q}^2 = varians kondisional bulan t-q

α_p = parameter ARCH, untuk p = 1,2,3,...,p

λ_q = parameter GARCH, untuk q=1,2,3,...,q

Tidak semua data runtun waktu mengandung efek heteroskedastisitas, diperlukan pengujian agar dapat mengetahui apakah ada tidaknya efek heteroskedastisitas dalam data runtun waktu.

Pengujian dapat dilakukan dengan cara memeriksa fungsi autokorelasi dari ε^2 (Wei W., 2006).

Langkah-langkah yang dapat ditempuh adalah:

1. Estimasi model untuk mean dari proses $\{r_t\}$, misal menggunakan model ARMA atau model regresi. Selanjutnya, hitung nilai fitted residual dari model ini $\varepsilon_t = r_t - \hat{\mu}_t$ dan $\hat{\varepsilon}_t^2$ dan selanjutnya hitunglah variansi dari $\hat{\varepsilon}_t^2$ sebagai berikut:

$$\sigma^2 = \frac{\sum_{t=1}^n \varepsilon_t^2}{n} \quad (10)$$

dimana n menyatakan banyaknya data dari residual.

2. Hitung dan buatlah plot sampel ACF dan PACF dari $\hat{\varepsilon}_t^2$. Fungsi sampel ACF dari $\hat{\varepsilon}_t^2$ dapat dihitung sebagai berikut:

$$\rho_k = \frac{\sum_{t=k+1}^T (\varepsilon_t^2 - \sigma^2)(\varepsilon_{t-k}^2 - \sigma^2)}{\sum_{t=1}^T (\varepsilon_t^2 - \sigma^2)^2} \quad (11)$$

Dan sampel PACF dihitung dengan menggunakan bentuk sampel PACF untuk proses autoregresi bagi $\hat{\varepsilon}_t^2$. Untuk sampel yang berukuran besar, sampel ACF dan PACF untuk $\hat{\varepsilon}_t^2$ dibandingkan dengan $\pm \frac{2}{\sqrt{n}}$. Apabila terdapat nilai sampel ACF dan PACF yang signifikan berbeda dengan nol (yakni nilainya berada di luar interval $\pm \frac{2}{\sqrt{n}}$ maka efek heteroskedastisitas dikatakan signifikan.

ARCH-GARCH

Atas dasar tersebut terdapat langkah-langkah untuk menentukan model ARCH-GARCH. Maka tahapan pemodelan metode ARCH-GARCH pada

data indeks harga saham PT Verena Multi Finance Tbk dari periode penelitian bulan Januari 2009 sampai dengan bulan Desember 2019, terdiri atas beberapa tahapan sebagai berikut.

1. Pengujian Efek Heteroskedastisitas

Apabila model ARIMA telah didapatkan, maka dapat dilakukan pengujian untuk mengetahui keberadaan efek heteroskedastisitas. Untuk menguji efek heteroskedastisitas pada data dilakukan pengujian pada residual kuadrat dari model terbaik yang telah diperoleh sebelumnya, yaitu dengan memakai uji Q-Ljung Box (Box & Pierce, 1970). Hipotesis sebagai berikut:

- Hipotesis penelitian
 $H_0 : (1) = \rho(2) = \dots = \rho(k) = 0$
 (artinya tidak terdapat efek heteroskedastisitas)
 $H_1 : \text{paling tidak terdapat satu } k \in 1,2,\dots,m$
 dengan $\rho(k) \neq 0$
 (artinya terdapat efek heteroskedastisitas)

- Statistik uji

$$Q = (n + 2) \sum_{i=1}^m \frac{p_i}{n - i} \quad (12)$$

dengan i = banyak lag

- Kriteria uji
 Tolak H_0 apabila $Q > \chi(i)^2$ atau p-value $< \alpha$ yang berarti dalam residual kuadrat terdapat ARCH/GARCH, dan terima H_0 dalam hal lainnya.

2. Penaksiran Parameter GARCH

Setelah melakukan pengujian keberadaan efek heteroskedastisitas, lihat plot ACF dan PACF dari residual kuadrat untuk pembentukan model ARCH-GARCH. Langkah selanjutnya adalah dari beberapa model yang telah didapatkan akan dilakukan penaksiran parameter model menggunakan maximum likelihood. Penaksiran parameter ini dilaksanakan untuk menguji keberartian parameter dalam model.

3. Pemilihan Model GARCH Terbaik

Setelah melakukan penaksiran parameter pada model GARCH, maka kita bisa tentukan model terbaiknya. Pilihlah model yang mengecilkan nilai kriteria seperti *Akaike Information Criteria* (AIC). (Nachrowi D., 2006)

4. Akurasi Peramalan

Model peramalan yang didapatkan akan divalidasi agar dapat mengetahui akurasi metode peramalan dalam pemodelan data time seriesnya. Salah satunya adalah dengan menghitung rata-rata penyimpangan absolut atau MAD. MAD menunjukkan adanya *error* pada data. Hal ini dapat diperoleh dengan mencari rata-rata nilai absolut error dari semua hasil *forecasting*. Nilai absolut

disini digunakan untuk menghindari nilai *error* yang positif serta nilai *error* yang negatif saling meniadakan. Persamaannya dapat ditulis seperti di bawah ini.

$$MAD = \frac{1}{n} \sum_{t=1}^m |X_t - F_t| \quad (13)$$

Dengan,

X : data aktual observasi

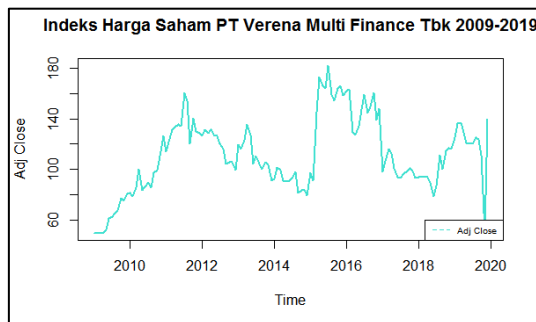
F : data peramalan.

3. HASIL DAN PEMBAHASAN ARIMA

Dengan menggunakan *software* R, akan dilakukan analisis dengan tahapan sebagai berikut.

1. Stasioneritas Data

Hal paling awal yang seharusnya dilakukan dalam mengecek kestasioneritasan data yaitu dengan membuat plot data deret waktu pada data indeks harga PT Verena Multi Finance Tbk dari periode Januari 2009 sampai Desember 2019. Pada plot data akan dilihat kestasioneritasan data dalam rerata dan varians.



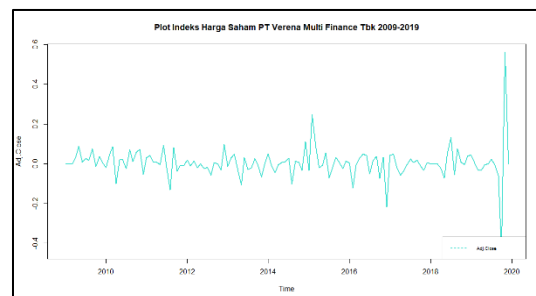
Gambar 1. Plot indeks harga saham PT Verena Multi Finance Tbk Januari 2009 – Desember 2019

Berdasarkan Gambar 1. terlihat bahwa data indeks harga saham PT Verena Multi Finance Tbk secara visual menunjukkan adanya pola *trend* pada selang waktu tertentu, seperti pada tahun 2009 hingga 2011 terlihat pola *trend* naik. Namun, setelah tahun 2011 hingga tahun 2015 terlihat pola *trend* menurun. Secara keseluruhan, plot data pada Gambar 1. tidak memiliki pola musiman dan pola stasioner karena data tidak menyebar di sekitaran rerata.

Kestasioneritasan dalam varians pada data dapat diperiksa menggunakan uji *Box-Cox*. Nilai lambda yang didapatkan setelah dilakukan uji *Box-Cox* terhadap data indeks harga saham PT Verena Multi

Finance Tbk adalah sebesar 0.1362227. dalam hal ini, data tidak memiliki sifat kestasioneritasan dalam varians, maka mesti ditransformasi. Sesudah dua kali ditransformasi, ternyata data memiliki nilai lambda yang meningkat hingga sebesar 0.7926443. Artinya, data sudah memenuhi salah satu syarat, yaitu stasioner dalam varians.

Pemeriksaan kestasioneritasan dalam rerata pada data dapat memakai Uji *Unit Root* dengan metode *Augmented Dicky-Fuller* (ADF) dengan hipotesis nolnya adalah data indeks harga saham PT Verena Multi Finance Tbk belum stasioner. Sedangkan untuk hipotesis alternatifnya adalah data indeks harga saham sudah stasioner. Berdasarkan pengujian ADF pada data indeks harga saham PT Verena Multi Finance Tbk diperoleh *p-value* = 0.1788. Maka, H_0 diterima artinya data indeks harga saham tidak stasioner. Oleh karena itu perlu dilakukan pembedaan (*differencing*). Sesudah data dilakukan *differencing* satu kali, didapatkan *p-value* sebesar 0.01, artinya $\alpha = 0.05$ lebih besar dari *p-value*. Sehingga, dapat disimpulkan bahwa data sudah memiliki kestasioneritasan dalam rerata.

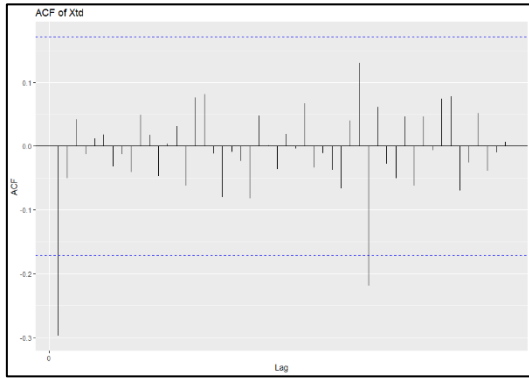


Gambar 2. Plot Indeks Harga Saham PT Verena Multi Finance Tbk Januari 2009 – Desember 2019 yang Sudah Stasioner

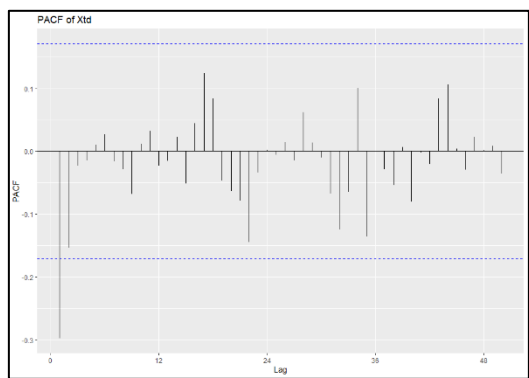
Pada gambar tersebut data dilihat bahwa plot data deret waktu indeks harga saham PT Verena Multi Finance Tbk sudah memiliki kestasioneritasan baik dalam varians maupun dalam rerata dengan *differencing* sebanyak sekali.

2. Identifikasi Model

Sesudah diuji asumsi kestasioneritasannya baik dalam varians maupun rata-rata terpenuhi, lalu tahapan berikutnya adalah penentuan orde model ARIMA yaitu $AR(p)$ dan $MA(q)$ yang didapatkan melalui penggambaran pola ACF dan PACF menurut data awal yang telah diuji kestasioneritasan pada varians dan rerata, sehingga dari plot ACF dan PACF bisa mengidentifikasi model ARIMA yang dibentuk.



Gambar 3. Plot ACF Data Indeks Harga Saham PT Verena Multi Finance Tbk Januari 2009 – Desember 2019



Gambar 4. Plot PACF Data Indeks Harga Saham PT Verena Multi Finance Tbk Januari 2009 – Desember 2019

Berdasarkan Gambar 3. dapat dilihat bahwa pola yang terbentuk dapat diduga beberapa nilai yaitu pada plot ACF dapat diidentifikasi orde MA melihat pada lag yang signifikan, sehingga diperoleh MA(0) dan MA(1). Berdasarkan Gambar 4. dapat dilihat pola PACF yang signifikan karena *cut off after lag 1* sehingga plot dapat diidentifikasi orde AR yaitu AR(0) dan AR(1). Selain itu, sorde untuk *differencing* adalah $d = 1$ karena data dilakukan *differencing* sekali. Beberapa alternatif model yang dapat digunakan adalah :

1. Model 1 : model ARIMA(1,1,0)
2. Model 2 : model ARIMA(0,1,1)
3. Model 3 : model ARIMA(1,1,1)

Sesudah didapatkan model dugaan ARIMA maka langkah berikutnya yaitu menaksir parameter untuk setiap model. Hasil penaksiran parameter tersebut akan diuji signifikansinya untuk melihat apakah parameter yang ditaksir memiliki keberartian atau pantas termasuk ke dalam suatu model. Untuk pengujian signifikansi parameter, memakai statistik uji t-test melalui kriteria signifikan apabila α lebih besar dari *p-value*.

- Hipotesis nolnya adalah parameter yang ditaksir tidak signifikan, sedangkan untuk hipotesis alternatifnya adalah parameter yang ditaksir sudah signifikan.
- Taraf Signifikan : $\alpha = 5\% = 0.05$
- Statistik Uji : Uji t-test

$$t_{hitung} = \frac{\theta}{se(\hat{\theta})} \tag{14}$$

3. Penaksiran dan Uji Signifikansi Parameter

Tabel 1. Hasil penaksiran parameter beserta uji signifikansi parameter

Model	Parameter	Estimasi	Standar Error	t-hitung	p-value	Keterangan
ARIMA(1,1,0)	AR(1)	-0.301009	0.098073	-3.0683	0.002146	Signifikan
ARIMA(0,1,1)	MA(1)	-0.311967	0.084173	-3.7063	0.0002103	Signifikan
ARIMA(1,1,1)	AR(1)	-0.02168	0.28078	-0.0772	0.9385	Tidak Signifikan
	MA(1)	-0.29392	0.24935	-1.1787	0.2385	Tidak Signifikan

4. Pemeriksaan Asumsi

Model yang terpenuhinya asumsi residual bersifat *white noise* atau residual yang tidak berautokorelasi adalah salah satu ciri model ARIMA yang baik. Agar mengetahui residual berautokorelasi atau tidak dilakukan pengujian dengan uji *Q-Ljung Box*. Apabila nilai α 0.05 lebih kecil dari *p-value* maka residual memenuhi syarat tidak

berautokorelasi. Hasil pengujian memakai uji *Q-Ljung Box* terlihat bahwa model ARIMA(1,1,0) beserta ARIMA(0,1,1) terpenuhi asumsi residual tidak berautokorelasi dengan nilai *p-value* berturut-turut untuk setiap model adalah 0.6955 dan 0.9517. Selanjutnya akan dilakukan pengujian kenormalitasan residual memakai uji Kolmogorov-Smirnov. Apabila nilai α 0.05 lebih kecil dari *p-value* maka asumsi normalitas residual terpenuhi.

Fajriyah Fitriyani¹, Salsabila Fasya A², Muhammad Irfan R³, Teguh Ammar T⁴/

Hasil pengujian memakai uji Kolmogorov-Smirnov memperlihatkan jika pada model ARIMA(1,1,0) beserta model ARIMA(0,1,1) memenuhi dugaan kenormalitasan residual dengan nilai *p-value* berturut-turut untuk setiap model adalah 0.05112 dan 0.07081.

5. Pemilihan Model ARIMA Terbaik

Berdasarkan dua model dugaan ARIMA yang dibentuk sebelumnya, semua parameter model signifikan, dan begitu pula untuk asumsi *white noise* dan normalitas residual, maka asumsi tersebut terpenuhi oleh kedua model tersebut. Untuk memilih model dugaan ARIMA terbaik, dapat dilihat dari model yang memiliki nilai AIC terkecil dari kedua model ARIMA tersebut. Berikut ini adalah nilai AIC untuk setiap dugaan model.

Tabel 2. Nilai AIC masing-masing model ARIMA indeks harga saham PT Verena Multi Finance Tbk

No	Model ARIMA	Nilai AIC
1	ARIMA(1,1,0)	1079.001
2	ARIMA(0,1,1)	1077.736

Jika dilihat dari tabel diatas, model yang mempunyai nilai AIC terkecil ialah model ARIMA(0,1,1). Maka, dapat diambil kesimpulan bahwa model ARIMA yang terpilih adalah model ARIMA(0,1,1). Untuk mendapatkan persamaan dari model ARIMA(0,1,1), substitusikan nilai estimasi parameter pada Tabel 1. ke dalam persamaan berikut.

$$X_t = \mu' + \phi X_{t-1} + \varepsilon_t \tag{15}$$

Sehingga akan diperoleh persamaan sebagai berikut.

$$X_t = -0.3120X_{t-1} + \varepsilon_t \tag{16}$$

ARCH-GARCH

1. Pengujian Efek Heteroskedastisitas

Sesudah diperoleh model ARIMA terbaik, tahap berikutnya yaitu pengujian efek heteroskedastisitas terhadap residual kuadrat model terbaik ialah model ARIMA(0,1,1). Pengujian efek heteroskedastisitas dilakukan dengan menggunakan uji *Q-Ljung Box*. Apabila *alpha* 0.05 lebih besar dari *p-value* artinya ada efek heteroskedastisitas. Pada hasil uji *Q-Ljung Box* diperoleh hasil nilai *p-value* yang lebih kecil dari *alpha* 0.05 yaitu sebesar 4.031e-06. Maka, dapat diambil kesimpulan bahwa ada efek heteroskedastisitas pada model ARIMA(0,1,1).

2. Pengujian Efek ARCH pada Model ARIMA

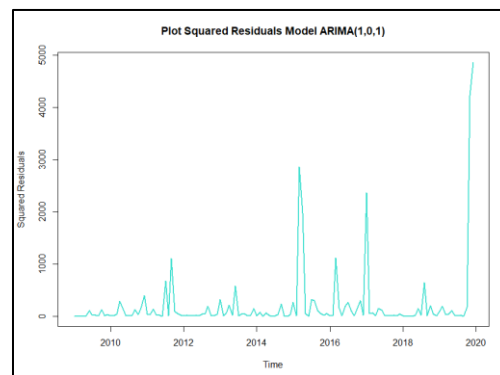
Fajriyah Fitriyani¹, Salsabila Fasya A², Muhammad Irfan R³, Teguh Ammar T⁴/
 J Statistika Vol. 14, No. 1, (2021)

Pengujian efek ARCH dalam residual persamaan kuadrat terkecil tunggal dapat memakai uji *Langrange Multiplier* (LM) dengan dugaan yaitu : Hipotesis nol berasumsi bahwa tidak ada efek ARCH pada model ARIMA, sedangkan hipotesis alternatif berkata sebaliknya, yaitu ada efek ARCH pada model ARIMA.

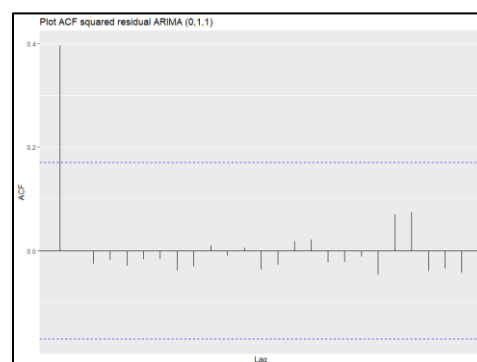
Apabila nilai *alpha* 0.05 lebih besar dari *p-value* maka residual memiliki efek ARCH pada model ARIMA(0,1,1). Hasil pengujian menggunakan uji *Langrange Multiplier* (LM) nampak bahwa pada model ARIMA(0,1,1) terdapat efek ARCH karena hasil perhitungan nilai *p-value* sebesar kurang dari 2.2e-16. Maka, setelah itu dilanjutkan dengan tahap mengidentifikasi model ARCH-GARCH.

3. Identifikasi Model ARCH-GARCH

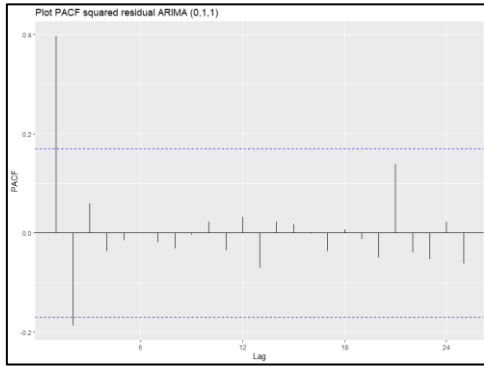
Pada tahap ini mengidentifikasi model ARCH(*p*) dan GARCH(*p,q*) mengikuti plot ACF beserta plot PACF residual kuadrat model ARIMA(0,1,1). Adapun plot ACF beserta plot PACF residual kuadrat model ARIMA(0,1,1) adalah sebagai berikut.



Gambar 5. Plot Residual Kuadrat Model ARIMA(0,1,1)



Gambar 6. Plot ACF Residual Kuadrat Model ARIMA(0,1,1)



Gambar 7. Plot PACF Residual Kuadrat Model ARIMA(0,1,1)

Pada gambar diatas terlihat bahwa plot ACF *cut off* pada lag pertama. Namun, untuk plot PACF terlihat *cut off after* lag ke-2. Sehingga, dapat dibentuk dua model ARCH(p) dan dua model GARCH(p,q). Beberapa alternatif model dugaan ARCH dan GARCH yang dapat digunakan :

1. Model 1 : ARCH(1)
2. Model 2 : ARCH(2)
3. Model 3 : GARCH(1,1)
4. Model 4 : GARCH(1,2)

4. Penaksiran beserta Uji Signifikansi Parameter ARCH-GARCH

Sesudah didapatkan model dugaan ARCH dan GARCH, maka tahap berikutnya yaitu penaksiran parameter untuk setiap model. Hasil penaksiran parameter tersebut akan diuji signifikansinya untuk melihat apakah parameter yang ditaksir memiliki keberartian atau layak masuk ke dalam suatu model. Hasil penaksiran parameter bersama uji kesignifikasian parameter dapat dilihat dalam tabel berikut ini.

Tabel 3. Hasil penaksiran parameter beserta uji signifikansi parameter model ARCH-GARCH

Model	Parameter	Estimasi	Standar Error	t-hitung	p-value	Keterangan
ARCH(1)	α_0	155.15429	11.48753	13.506	<2e-16	Signifikan
	α_1	0.21975	0.09237	2.379	0.0174	Signifikan
ARCH(2)	α_0	95.1986	15.1059	6.302	2.94e-10	Signifikan
	α_1	0.4237	0.1182	3.584	0.000338	Signifikan
	α_2	0.4750	0.1183	4.016	5.92e-05	Signifikan
GARCH(1,1)	α_0	1.905e+02	1.049e+02	1.816	0.0694	Tidak Signifikan
	α_1	1.624e-01	7.802e-02	2.082	0.0373	Signifikan
	β_1	3.426e-12	4.719e-01	0	1	Tidak Signifikan
GARCH(1,2)	α_0	1.799e+02	3.848e+02	0.468	0.640	Tidak Signifikan
	α_1	1.672e-01	8.097e-02	2.064	0.039	Signifikan
	α_2	4.843e-02	3.807e-01	0.127	0.899	Tidak Signifikan
	β_1	5.435e-12	2.119	0	1	Tidak Signifikan

Dari hasil di atas terlihat bahwa model ARCH(1) memiliki parameter α_0 dan α_1 yang memiliki keberartian dalam model. Begitu pula dengan model ARCH(2) memiliki parameter α_0 , α_1 , dan α_2 yang memiliki keberartian juga dalam model. Model GARCH(1,1) memiliki parameter α_1 yang berarti dalam model, tetapi tidak dengan parameter α_0 dan β_1 . Model GARCH(1,2) memiliki parameter α_1 yang berarti dalam model, tetapi tidak dengan parameter α_0 , α_2 , dan β_1 .

5. Pemilihan Model ARCH-GARCH Terbaik

Berdasarkan model dugaan ARCH dan GARCH yang telah dibentuk sebelumnya, selanjutnya akan dipilih model dugaan terbaik yang akan digunakan untuk menduga indeks harga saham PT Verena

Multi Finance Tbk. Untuk memilih model dugaan ARCH-GARCH terbaik, dapat dilihat dari model dengan nilai AIC terkecil. Untuk menentukan orde ARCH/GARCH yang akan digunakan, akan dihitung AIC dengan rumus sebagai berikut :

$$AIC_c = -2 \log likelihood + 2(q + 1) \left(\frac{N}{N-q-2} \right), \tag{17}$$

tidak konstan.

$$AIC_c = -2 \log likelihood + 2(q + 2) \left(\frac{N}{N-q-3} \right), \tag{18}$$

konstan.

Dengan :

N = Ukuran Data

q = Orde Autoregressive

Fajriyah Fitriyani¹, Salsabila Fasya A², Muhammad Irfan R³, Teguh Ammar T⁴/

Tabel 4. Nilai AIC Masing-Masing Model ARCH-GARCH Indeks Harga Saham PT Verena Multi Finance Tbk

Model	N	Q	Log Likelihood	AICc Tidak Konstan	AICc Konstan
ARCH(1)	132	1	-526.0697	1056.233	1058.327
ARCH(2)	132	2	-519.5263	1045.24	1047.368

Dari Tabel 4. terlihat AIC baik konstan maupun non konstan semakin menurun, dan pada model ARCH(2) memiliki AIC terkecil, sehingga model ARCH-GARCH yang akan dipakai untuk menduga indeks harga saham PT Verena Multi Finance Tbk adalah ARCH(2).

Untuk mendapatkan persamaan dari model ARCH(2), substitusikan taksiran parameter pada Tabel 3. ke dalam persamaan berikut.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 \quad (19)$$

Sehingga, akan diperoleh persamaan sebagai berikut.

$$\sigma_t^2 = 95.1986 + 0.4237 \varepsilon_{t-1}^2 + 0.475 \varepsilon_{t-2}^2 \quad (20)$$

Persamaan tersebut memiliki arti bahwa variansi indeks harga saham PT Verena Multi Finance Tbk pada periode t ditemukan oleh suatu konstanta (95.1986), sisaan kuadrat pada satu periode sebelumnya dengan proporsi sebesar 42.37%, dan sisaan kuadrat pada dua periode sebelumnya dengan proporsi sebesar 47.5%.

6. Peramalan Indeks Harga Saham PT Verena Multi Finance Tbk

Setelah memilih model ARCH-GARCH terbaik, maka akan diperoleh model gabungan ARIMA(0,1,1)-ARCH(2) untuk memprediksi indeks harga saham PT Verena Multi Finance Tbk sebagai berikut.

$$X_t = -0.3120X_{t-1} + \varepsilon_t + 95.1986 + 0.4237 \varepsilon_{t-1}^2 + 0.475 \varepsilon_{t-2}^2 \quad (21)$$

Berikutnya akan dilakukan peramalan (*forecasting*) menggunakan model ARIMA(0,1,1) dan variansi residual ARCH(2). Hasil peramalan indeks harga saham PT Verena Multi Finance Tbk untuk dua tahun kedepan diperlihatkan pada tabel di bawah ini.

Tabel 5. Hasil peramalan indeks harga saham PT Verena Multi Finance Tbk

Periode	Forecast	Sigma
1 Januari 2020	112.2	17.31
1 Februari 2020	102.4	16.77

1 Maret 2020	102.4	17.82
1 April 2020	102.4	17.97
1 Mei 2020	102.4	18.46
1 Juni 2020	102.4	18.71
1 Juli 2020	102.4	19.01
1 Agustus 2020	102.4	19.22
1 September 2020	102.4	19.42
1 Oktober 2020	102.4	19.59
1 November 2020	102.4	19.74
1 Desember 2020	102.4	19.87
1 Januari 2021	102.4	19.98
1 Februari 2021	102.4	20.08
1 Maret 2021	102.4	20.16
1 April 2021	102.4	20.23
1 Mei 2021	102.4	20.30
1 Juni 2021	102.4	20.35
1 Juli 2021	102.4	20.40
1 Agustus 2021	102.4	20.44
1 September 2021	102.4	20.48
1 Oktober 2021	102.4	20.51
1 November 2021	102.4	20.54
1 Desember 2021	102.4	20.56

Langkah berikutnya akan dilaksanakan evaluasi ketepatan model pendugaan antara persamaan model ARIMA(0,1,1)-ARCH(2) yang telah dibentuk dengan data indeks harga saham PT Verena Multi Finance Tbk. Evaluasi tersebut menggunakan MAD (*Mean Absolute Deviation*) sebagai ukurannya. Hasil prediksi tidak terpaut jauh jika dibandingkan oleh data asli, artinya model yang terbentuk sudah cukup baik untuk memprediksi indeks harga saham PT Verena Multi Finance Tbk. Nilai MAD hasil peramalan indeks harga saham PT Verena Multi Finance Tbk sebesar 28.01674. Hasil perhitungan MAD membuktikan bahwa model yang didapat sudah lumayan baik digunakan dalam memprediksi. Selain itu, evaluasi juga dilaksanakan dengan menghitung nilai MAPE (*Mean Absolute Percentage Error*). Nilai MAPE

Fajriyah Fitriyani¹, Salsabila Fasya A², Muhammad Irfan R³, Teguh Ammar T⁴/

hasil peramalan indeks harga saham PT Verena Multi Finance Tbk sebesar 22.9971%.

Menurut Lewis (1982), nilai MAPE dapat digolongkan ke dalam 4 kriteria yaitu jika MAPE bernilai kurang dari 10% maka hasil prediksi digolongkan ke dalam sangat akurat. Untuk MAPE bernilai antara 10% dan 20% maka hasil prediksi dikatakan baik. Nilai MAPE sebesar 20% hingga 50% dikategorikan wajar, dan untuk nilai MAPE lebih besar dari 50% maka digolongkan ke dalam tidak akurat.

Maka dari penggolongan tersebut terlihat bahwa nilai MAPE berbanding lurus dengan kesalahan hasil prediksi. Berdasarkan Lewis (1982), maka hasil peramalan indeks harga saham PT Verena Multi Finance Tbk memiliki kemampuan pendugaan yang wajar.

4. KESIMPULAN

Jika dilihat dari hasil analisis pada bagian hasil dan pembahasan, dapat diambil beberapa kesimpulan, yaitu :

1. Pada kasus data indeks harga saham PT Verena Multi Finance Tbk menunjukkan ketidakstabilan variansi residual yaitu terdapat efek heteroskedastisitas. Model ARIMA terbaik yaitu ARIMA(0,1,1) tidak dapat digunakan dalam mengatasi data yang mengandung efek heteroskedastisitas, maka digunakanlah model GARCH. Model GARCH terbaik yang diperoleh untuk melakukan peramalan indeks harga saham PT Verena Multi Finance Tbk adalah model ARCH(2). Maka, didapatkan persamaan model ARIMA(0,1,1)-ARCH(2) seperti dibawah ini.

$$X_t = -0.3120X_{t-1} + \varepsilon_t + 95.1986 + 0.4237\varepsilon_{t-1} + 0.475\varepsilon_{t-2}$$

2. Model ARIMA(0,1,1)-ARCH(2) mampu memprediksi secara baik data indeks harga saham PT Verena Multi Finance Tbk, hal ini dapat dilihat dari akurasi peramalan memakai MAD (Mean Absolute Deviation) dan MAPE (Mean Absolute Percentage Error). Didapatkan nilai MAD sebesar 28.01674 dan nilai MAPE sebesar 22.9971%. Dari nilai MAD dan MAPE tersebut terlihat jika model yang didapat sudah lumayan baik dalam memprediksi.

DAFTAR PUSTAKA

- Arch, M., Garch, D. A. N., Peramalan, P., Price, S., Corp, D., Bilondatu, R. N., & Isa, D. R. (2019). *HARGA SAHAM PT. COWELL DEVELOPMENT Tbk*. 13, 9–18.
- Ari Pani Desvina, N. R. (2016). Penerapan Metode ARCH / GARCH dalam Peramalan Indeks

- Harga Saham Sektor. *Jurnal Sains Matematika Dan Statistika*, 2(I), 1–10.
- Bollerslev, T. (1986). Generalized Auto Regressive Heteroscedastic Model, *Journal of Econometric*, 31 : 307 - 327.
- Faustina, R. S., Agoestanto, A., & Hendikawati, P. (2017). Model Hybrid ARIMA-GARCH untuk Estimasi Volatilitas Harga Emas. *UNNES Journal of Mathematics*, 6(1), 11–24.
- Garch-m, H. I. N. M. (2014). *Peramalan volatilitas menggunakan model*. 3, 655–662.
- Gaussian, J. (2016). 1 1 , 2 , 3. 5, 705–715.
- Irawan, W. (2019). Peramalan Harga Saham PT. Unilever Tbk dengan Menggunakan Metode ARIMA. *Jurnal Matematika UNAND*, 4(3), 80. <https://doi.org/10.25077/jmu.4.3.80-89.2015>
- Islam, U., Sunan, N., Yogyakarta, K., Islam, U., & Sunan, N. (n.d.). *GABUNGAN REGIONAL ASIA TERHADAP INDEKS HARGA SAHAM GABUNGAN INDONESIA* Widodo. 4988, 148–164.
- Juanda, Bambang, & Junaidi. (2012). *Ekonometrika Deret Waktu Teori dan Aplikasi*, Bogor: IPB Press.
- Listya, D., & Sutijo, B. (2013). Metode Peramalan Indeks Harga Saham Gabungan (IHSG) Nikkei 255 dengan Pendekatan Fungsi Transfer. *Jurnal Sains Dan Seni Pomits*, 2(2), D-271-D-274.
- Matematika, J., Matematika, F., Ilmu, D. A. N., & Alam, P. (2018). *Analisis peramalan indeks harga konsumen menggunakan metode arch garch berbantuan software eviews*.
- Metode, M., & Dengan, G. (2016). Analisis Volatility Forecasting Sembilan Bahan Pokok Menggunakan Metode Garch Dengan Program R. *Unnes Journal of Mathematics*, 5(1), 90–99. <https://doi.org/10.15294/ujm.v5i1.13109>
- Puspatika, K., & Kusumawati, Y. (2018). Peramalan Harga Cabai Dengan Metode Arima Arch- Garch Dan Single Moving Average Di Kota Semarang. *Journal JOINS Udinus*, 03(02), 192–201.
- Saida, M. D. N., Sudarno, & Hoyyi, A. (2016). Pemodelan Return Indeks Harga Saham Gabungan Menggunakan Threshold Generalized Autoregressive Conditional Heteroscedasticity (Tgarch). *Jurnal Gaussian*, 5(3), 465–474.
- Sari, F. T., Mariani, S., Matematika, J., & Semarang, U. N. (2016). Perbandingan Taksiran Value at Risk Dengan Program R

- Dan Matlab Analisis Investasi Saham Menggunakan Metode Garch. *Unnes Journal of Mathematics*, 5(2), 118–126. <https://doi.org/10.15294/ujm.v5i2.13120>
- Sidiq, A., Indeks, P., Sti, S., & Pergerakan, H. T. (2010). *INDEKS SAHAM GABUNGAN PADA BEI Ahmad Sidiq. 1*(November), 1–18.
- Syukrina, Filzah. (2020). *Penerapan Metode Peramalan Model GARCH Dalam Memprediksi Indeks Dst.* Skripsi. Tidak diterbitkan. Jatinangor: Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Padjadjaran.
- Tita Deitiana1; Stella. (2009). *Harga Saham Gabungan Bursa Efek Indonesia.* 10(1), 22–30.
- Zach. (2020). How to Calculate Mean Absolute Percentage Error (MAPE) in Excel. <https://www.statology.org/mape-excel/> (diakses tanggal 15 Mei 2021)

Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Provinsi Jawa Barat dengan Pendekatan Regresi Nonparametrik *Spline Truncated*Andrea Tri Rian Dani ⁽¹⁾, Ludia Ni'matuzzahroh ⁽²⁾

Institut Teknologi Sepuluh Nopember

Jl. Arif Rahman Hakim, Kota Surabaya 60111, Jawa Timur, Indonesia

e-mail: andreatriandani98@gmail.com, ludianimatuzzahroh@gmail.com

ABSTRAK

Estimator *Spline Truncated* adalah bentuk dari polinomial yang tersegmentasi dan bersifat kontinu, serta memiliki fleksibilitas yang tinggi dalam proses pemodelan. Estimator ini termasuk ke dalam pendekatan regresi nonparametrik yang dapat digunakan ketika bentuk hubungan antara variabel respon dan variabel prediktor menunjukkan pola yang berubah-ubah pada sub-sub interval tertentu. Penelitian ini bertujuan untuk memodelkan persentase penduduk miskin Kabupaten/Kota di Provinsi Jawa Barat dengan menggunakan model regresi nonparametrik estimator *Spline Truncated*. Metode estimasi yang digunakan adalah *Ordinary Least Squares* (OLS). Kriteria kebaikan model yang digunakan adalah *Generalized Cross-Validation* (GCV). Berdasarkan hasil analisis, didapatkan model terbaik regresi nonparametrik *Spline Truncated*, yaitu model dengan 3 titik knot, dimana diperoleh nilai GCV minimum sebesar 2.14. Berdasarkan hasil pengujian hipotesis, baik secara simultan maupun parsial, diketahui bahwa variabel prediktor yang digunakan pada penelitian ini, berpengaruh signifikan terhadap persentase penduduk miskin, dengan nilai koefisien determinasi sebesar 95.33%.

Kata kunci : *Generalized Cross-Validation (GCV), Persentase Penduduk Miskin, Spline Truncated.*

ABSTRACT

The Truncated Spline Estimators are polynomial pieces that have segmented and continuous properties and have high flexibility in the modeling process. The truncated Spline is one approach in nonparametric regression that can be used when the pattern of the relationship between the response variable and the predictor variable varies in certain sub-intervals. The purpose of this study is to model the percentage of poor people in districts/cities in West Java Province using a nonparametric regression model of the Spline Truncated estimator. The estimation method used is Ordinary Least Squares (OLS). The criteria for the goodness of the nonparametric regression model used is the Generalized Cross-Validation (GCV). Based on the analysis, the best model is obtained from the Truncated Spline nonparametric regression, namely the model with 3-knot points, where the minimum GCV value is 2.14. Based on the results of hypothesis testing, both simultaneously and partially, it is known that the predictor variables used in this study have a significant effect on the percentage of poor people, with a coefficient of determination of 95.33%.

Keywords : *Generalized Cross-Validation (GCV), Percentage of Poor People, Truncated Spline.*

1. PENDAHULUAN

Analisis regresi merupakan salah satu metode statistik yang dapat digunakan untuk mengetahui bentuk pola hubungan antara variabel respon dan variabel prediktor (Draper & Smith, 1992), (Montgomery, dkk., 2012). Tujuan utama pemodelan menggunakan analisis regresi adalah menentukan bentuk dari estimasi kurva regresi yang tepat (Budiantara, 2019). Bentuk hubungan dari variabel respon (dependen) dengan variabel

prediktor (independen) dapat didekati dengan pendekatan regresi nonparametrik.

Regresi nonparametrik dapat digunakan pada saat bentuk kurva regresi memiliki pola yang tidak diketahui. Salah satu estimator yang paling banyak digunakan dalam pemodelan dengan pendekatan regresi nonparametrik adalah estimator *Spline Truncated*. Hal ini dikarenakan pendekatan dengan model regresi nonparametrik yaitu *Spline Truncated* mempunyai interpretasi statistik dan

visual yang spesial dalam prakteknya dan sangat baik (Eubank, 1999). Selain itu, estimator *Spline Truncated* memiliki keunggulan karena mampu menangani perubahan data yang berubah-ubah pada interval-interval tertentu (Budiantara, 2009); (Dani & Adrianingsih, 2021). Dalam membangun model regresi menggunakan estimator *Spline Truncated*, ada beberapa hal yang perlu diperhatikan, diantaranya menentukan derajat dari model regresi, jumlah titik knot, dan lokasi titik knot berada (Montoya, dkk., 2014). Dengan menggunakan titik knot, kurva regresi dengan *Spline Truncated* mampu menangani problematika perubahan pola data yang tajam, sehingga memungkinkan untuk diperoleh hasil estimasi kurva yang relatif mulus (Dani, dkk., 2021).

Kelebihan dari estimator *Spline Truncated* itu sendiri yaitu memberikan perhitungan matematis yang lebih mudah dan sederhana. Adapun beberapa penelitian sebelumnya yang menggunakan pendekatan regresi nonparametrik dengan estimator *Spline Truncated*, diantaranya Yanthi dan Budiantara (2016); Saputro, dkk. (2018); Rahim, dkk. (2019); Sholikha, dkk. (2019); dan Dani, dkk. (2021).

Dalam analisis statistika, selain melakukan analisis statistika deskriptif juga perlu dilakukan analisis statistika inferensia, seperti pengujian hipotesis, baik secara simultan dan parsial terhadap parameter model untuk mengetahui apakah parameter tersebut signifikan terhadap model atau tidak. Beberapa penelitian mengenai pengujian hipotesis dalam regresi nonparametrik dengan estimator *Spline Truncated*, diantaranya Bidari dan Budiantara (2020) dan Dani, dkk. (2020).

Model regresi nonparametrik dapat diterapkan dalam berbagai disiplin ilmu seperti sosial, ekonomi, dan kedokteran. Salah satu kajian pada bidang sosial yang dapat diterapkan dalam pemodelan regresi nonparametrik dengan estimator *Spline Truncated* adalah masalah kemiskinan. Kemiskinan merupakan salah satu masalah yang serius yang dihadapi Indonesia, khususnya pada Provinsi Jawa Barat. Berdasarkan data BPS pada Bulan Maret 2020, di Provinsi Jawa Barat besarnya persentase penduduk miskin yaitu 7.88%, kemudian pada Bulan September 2020 menjadi sebesar 8.43%, sehingga dalam waktu satu tahun mengalami peningkatan sebesar 0.55%.

Berdasarkan paparan di atas, maka pada penelitian ini akan digunakan pendekatan regresi nonparametrik dengan estimator *Spline Truncated* dalam memodelkan persentase penduduk miskin

Kabupaten/Kota di Provinsi Jawa Barat pada Tahun 2020, kemudian akan dilakukan analisis statistika inferensia yaitu pengujian hipotesis baik secara simultan maupun parsial terhadap parameter model regresi yang dihasilkan. Manfaat yang diharapkan dari hasil penelitian ini yaitu adalah menambah wawasan keilmuan mengenai regresi nonparametrik *Spline Truncated* dan dapat dijadikan masukan untuk pemerintah daerah dan pemerintah pusat dalam menentukan kebijakan pemberantasan kemiskinan.

2. METODE PENELITIAN

Bagian metodologi penelitian ini menjelaskan tentang sumber data, variabel penelitian, serta langkah-langkah analisis data yang akan digunakan untuk mengatasi masalah Persentase Penduduk Miskin Kabupaten/Kota di Provinsi Jawa Barat Tahun 2020.

2.1 Data Penelitian

Penelitian ini menggunakan data sekunder yang dipublikasikan oleh Badan Pusat Statistik (BPS). Unit observasi yang digunakan adalah Kabupaten/Kota di Provinsi Jawa Barat Tahun 2020 sebanyak 25 Kabupaten/Kota.

Variabel yang digunakan terdiri dari satu variabel respon (Y) dan empat variabel prediktor (X). Variabel-variabel penelitian telah disajikan pada Tabel 1 sebagai berikut:

Tabel 1. Variabel Penelitian

Variabel	Simbol Variabel	Deskripsi
Respon	Y	Persentase Penduduk Miskin
	X_1	Angka Harapan Hidup
	X_2	Rata-Rata Lama Sekolah
Prediktor	X_3	Gini Ratio
	X_4	Tingkat Pengangguran Terbuka

2.2 Tahapan Analisis Data

Berikut ini merupakan tahapan-tahapan analisis dalam pemodelan regresi nonparametrik dengan estimator *Spline Truncated*:

1. Melakukan identifikasi setiap variabel penelitian menggunakan statistika deskriptif.
2. Membentuk *scatter plot* antara variabel respon dengan setiap variabel prediktor.
3. Melakukan deteksi multikolinearitas pada setiap variabel prediktor.
4. Memodelkan data Persentase Penduduk Miskin Kabupaten/Kota di Provinsi Jawa Barat Tahun 2020 dengan model regresi nonparametrik *Spline Truncated*.

5. Memilih titik knot optimal yang menghasilkan nilai *Generalized Cross-Validation* (GCV) minimum.
6. Menghitung nilai koefisien determinasi (R^2).
7. Melakukan pengujian hipotesis secara simultan (Uji F).
8. Melakukan pengujian hipotesis secara parsial (Uji t).
9. Pengujian asumsi klasik yaitu, asumsi residual identik, asumsi residual independen, dan asumsi residual berdistribusi normal.

3. HASIL DAN PEMBAHASAN

Pada bagian hasil dan pembahasan ini akan dijelaskan mengenai pemodelan persentase penduduk miskin Kabupaten/Kota di Provinsi Jawa Barat dengan Regresi Nonparametrik *Spline Truncated*.

3.1 Statistika Deskriptif

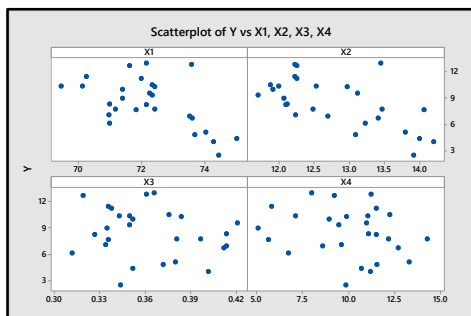
Berikut akan ditampilkan statistika deskriptif yang dapat memberikan informasi awal tentang masing-masing variabel yang digunakan, baik variabel respon maupun variabel prediktor pada Tabel 2.

Tabel 2. Statistika Deskriptif

Variabel	Minimum	Maksimum	Rata-Rata	Standar Deviasi
Y	2.45	12.97	8.42	2.80
X ₁	69.47	75.01	72.24	1.40
X ₂	11.7	14.2	12.79	0.77
X ₃	0.31	0.42	0.36	0.03
X ₄	5.08	14.29	9.98	2.38

3.2 Diagram Pencar

Bentuk pola hubungan antar variabel respon dengan setiap variabel prediktor, dapat dilihat dengan menggunakan bantuan diagram pencar (*scatter plot*) yang ditampilkan pada Gambar 1.



Gambar 1. Diagram Pencar

Berdasarkan Gambar 1, dapat diketahui bahwa pola hubungan antara variabel respon dengan setiap variabel prediktor terlihat tidak mengikuti bentuk pola tertentu, yang selanjutnya

diduga memiliki pola yang berubah pada sub-sub interval tertentu, sehingga akan digunakan pendekatan regresi nonparametrik dengan estimator *Spline Truncated*.

3.3 Deteksi Multikolinearitas

Pada kasus pemodelan menggunakan analisis regresi, tidak diperbolehkan apabila terdapat masalah multikolinearitas pada setiap variabel prediktor. Akan digunakan nilai *Variance Inflation Factor* (VIF) untuk mendeteksi masalah multikolinearitas.

Tabel 3. Nilai VIF

Variabel	Variance Inflation Factor (VIF)
X ₁	2.01
X ₂	1.72
X ₃	1.46
X ₄	1.62

Berdasarkan Tabel 3, dapat diketahui jika tidak terdapat nilai VIF yang lebih dari 10, sedemikian sehingga tidak terdapat masalah multikolinearitas pada variabel prediktor yang digunakan pada kasus ini.

3.4 Pemodelan Regresi Nonparametrik dengan Estimator *Spline Truncated*

1. Pemilihan Titik Knot Optimal

Pada bagian ini, akan dibahas mengenai proses pemilihan titik yang optimal. Langkah awal yang dilakukan adalah menentukan banyaknya titik knot serta lokasi dimana titik knot berada. Penelitian ini menggunakan jumlah titik knot dengan jumlah yang sama untuk setiap variabel prediktor, yaitu 1 sampai dengan 3 titik knot. Model regresi nonparametrik yang terbaik, akan dipilih dengan melihat nilai GCV terkecil dari beberapa model yang dicobakan.

Tabel 4. Nilai GCV

Banyaknya Knot	Lokasi Titik Knot				GCV
	X ₁	X ₂	X ₃	X ₄	
1	72.01	12.85	0.36	9.30	6.29
2	71.05	12.41	0.34	7.71	3.02
	71.58	12.65	0.35	8.59	
3	70.16	12.01	0.33	6.23	2.14
	70.62	12.22	0.34	6.99	
	71.55	12.64	0.35	8.54	

Berdasarkan Tabel 4, terlihat bahwa model regresi nonparametrik dengan estimator *Spline Truncated* terbaik adalah model dengan 3 titik knot, dimana diperoleh nilai GCV terkecil yaitu sebesar 2.14.

Adapun lokasi titik knot yang optimal untuk setiap variabel prediktor sebagai berikut:

Pada variabel X_1 (Angka Harapan Hidup)
 $K_1 = 70.16$ $K_2 = 70.62$ $K_3 = 71.55$

Pada variabel X_2 (Rata-Rata Lama Sekolah)
 $K_1 = 12.01$ $K_2 = 12.22$ $K_3 = 12.64$

Pada variabel X_3 (Gini Ratio)
 $K_1 = 0.33$ $K_2 = 0.34$ $K_3 = 0.35$

Pada variabel X_4 (Tingkat Pengangguran Terbuka)
 $K_1 = 6.23$ $K_2 = 6.99$ $K_3 = 8.54$

Selanjutnya akan dituliskan model regresi nonparametrik *Spline Truncated* terbaik:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 (X_1 - 70.16)_+ + \dots + \hat{\beta}_4 (X_1 - 71.55)_+ + \hat{\beta}_5 X_2 + \hat{\beta}_6 (X_2 - 12.01)_+ + \dots + \hat{\beta}_8 (X_2 - 12.64)_+ + \hat{\beta}_9 X_3 + \hat{\beta}_{10} (X_3 - 0.33)_+ + \dots + \hat{\beta}_{12} (X_3 - 0.35)_+ + \hat{\beta}_{13} X_4 + \hat{\beta}_{14} (X_4 - 6.23)_+ + \dots + \hat{\beta}_{16} (X_4 - 8.54)_+$$

Nilai Koefisien Determinasi (R^2) yang dihasilkan, yaitu sebesar 95.33%.

2. Pengujian Hipotesis Simultan

Berdasarkan dari hasil pemilihan titik knot yang optimal, maka langkah selanjutnya adalah melakukan estimasi dan pengujian parameter model secara simultan. Hipotesis yang digunakan pada pengujian hipotesis simultan sebagai berikut:

Hipotesis
 H_0 : $\beta_1 = \beta_2 = \dots = \beta_{16} = 0$
 H_1 : Minimal terdapat satu $\beta_j \neq 0$,
 $j = 1, 2, \dots, 16$

Dengan taraf signifikansi sebesar 5% dan Daerah Kritis yaitu Tolak H_0 jika $F_{hit} > F_{(\alpha, k, n-k-1)}$ maka:

Tabel 5. ANOVA

Sumber	df	SS	MS	Fhit
Regresi	16	195.43	12.21	34.44
Error	10	9.57	0.35	
Total	26	205.02		

Berdasarkan Tabel 5, diperoleh nilai F_{hit} sebesar 34.44 dimana lebih besar daripada nilai $F_{(0.05, 16, 10)}$ sebesar 2.83, sehingga diputuskan Tolak H_0 . Dapat disimpulkan jika minimal terdapat satu parameter yang signifikan terhadap variabel respon. Selanjutnya perlu dilakukan pengujian secara parsial guna mengetahui parameter-parameter yang memberikan pengaruh signifikan.

3. Pengujian Parameter secara Parsial

Pada tahapan ini, akan dilakukan pengujian signifikansi parameter secara parsial dengan hipotesis sebagai berikut:
 Hipotesis

H_0 : $\beta_j = 0$
 H_1 : $\beta_j \neq 0, j = 1, 2, \dots, 16$

Dengan taraf signifikansi sebesar 5%, maka:

Tabel 6. Estimasi Parameter dan Uji t

Variabel	Estimasi Parameter	P-Value	Kesimpulan	
X_1	$\hat{\beta}_0$	-515.88	1.08×10^{-2}	Signifikan
	$\hat{\beta}_1$	7.41	1.03×10^{-2}	Signifikan
	$\hat{\beta}_2$	-27.59	4.89×10^{-5}	Signifikan
	$\hat{\beta}_3$	31.47	7.06×10^{-7}	Signifikan
	$\hat{\beta}_4$	-13.34	6.59×10^{-8}	Signifikan
X_2	$\hat{\beta}_5$	3.94	2.39×10^{-1}	Tidak Signifikan
	$\hat{\beta}_6$	17.89	1.17×10^{-2}	Signifikan
	$\hat{\beta}_7$	-37.17	4.71×10^{-6}	Signifikan
	$\hat{\beta}_8$	15.30	1.06×10^{-5}	Signifikan
X_3	$\hat{\beta}_9$	-143.93	3.62×10^{-3}	Signifikan
	$\hat{\beta}_{10}$	262.65	1.77×10^{-4}	Signifikan
	$\hat{\beta}_{11}$	105.55	2.47×10^{-1}	Tidak Signifikan
	$\hat{\beta}_{12}$	-223.13	3.93×10^{-4}	Signifikan
X_4	$\hat{\beta}_{13}$	0.54	7.26×10^{-1}	Tidak Signifikan
	$\hat{\beta}_{14}$	6.59	1.03×10^{-1}	Tidak Signifikan
	$\hat{\beta}_{15}$	-11.35	8.42×10^{-3}	Signifikan
	$\hat{\beta}_{16}$	3.92	3.23×10^{-3}	Signifikan

Berdasarkan Tabel 6, diketahui bahwa terdapat beberapa parameter yang tidak signifikan. Dari 16 parameter yang ada pada model, terdapat 4 parameter yang tidak signifikan, dikarenakan memiliki *P-Value* yang lebih besar dari taraf signifikansi yang telah ditentukan yaitu sebesar 5% (0.05). Meskipun demikian, dapat disimpulkan secara garis besar bahwa semua variabel berpengaruh terhadap variabel respon yaitu persentase penduduk miskin Kabupaten/Kota di Jawa Barat.

3.5 Uji Asumsi

1. Asumsi Residual Identik

Pengujian asumsi residual identik dapat dilakukan untuk melihat residual dari hasil pemodelan memiliki varians yang sama (homogen) atau tidak. Adapun hipotesis yang digunakan sebagai berikut:

Hipotesis
 H_0 : $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$
 H_1 : Minimal terdapat satu $\sigma_i^2 \neq \sigma^2$, untuk
 $i = 1, 2, \dots, n$

Dengan menggunakan uji Glejser, didapatkan hasil sebagai berikut:

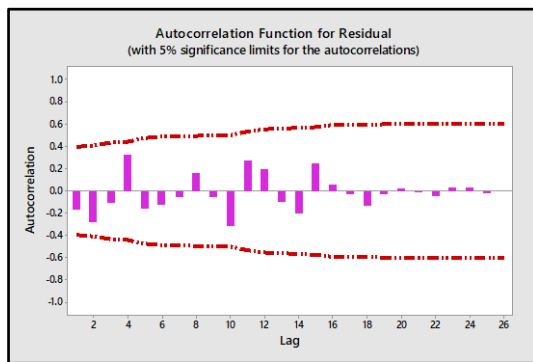
Tabel 7. Uji Glejser

Sumber	Df	SS	MS	Fhit
Regresi	16	2.09	0.13	0.71
Error	10	1.85	0.18	
Total	26	3.94		

Berdasarkan Tabel 7, diperoleh nilai F_{hit} sebesar 0.71 dimana lebih kecil dari nilai $F_{(0.05,16,10)}$ sebesar 2.83, sehingga diputuskan gagal Tolak H_0 . Dapat disimpulkan bahwa residual dari hasil pemodelan adalah homogen atau dapat dikatakan tidak terjadi masalah heteroskedastisitas.

2. Asumsi Residual Independen

Pengujian asumsi residual independen digunakan untuk melihat ataupun mendeteksi apakah ada otokorelasi antar residual dari hasil pemodelan. Pada penelitian ini, digunakan grafik ACF untuk mendeteksi masalah otokorelasi. Asumsi residual independen terpenuhi apabila tidak ada lag pada grafik ACF yang keluar dari batas signifikansi yang telah ditentukan. Ditampilkan grafik ACF dari residual pada Gambar 2.



Gambar 2. Grafik ACF dari Residual

Berdasarkan Gambar 2, diketahui jika tidak ada satupun lag yang keluar dari batas signifikansi, sehingga terbukti tidak terjadi masalah otokorelasi.

3. Asumsi Residual Berdistribusi Normal

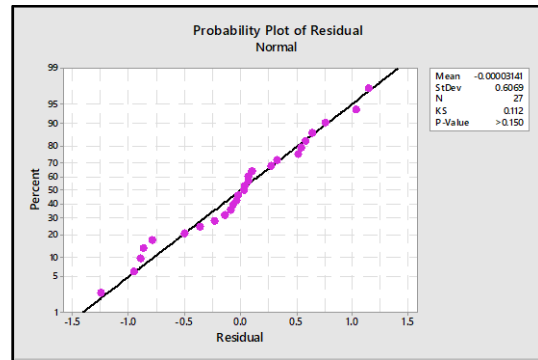
Statistik uji yang digunakan pada penelitian ini untuk mendeteksi residual berdistribusi normal adalah Kolmogorov-Smirnov. Adapun hipotesis yang digunakan sebagai berikut:

Hipotesis

H_0 : Residual berdistribusi Normal

H_1 : Residual tidak berdistribusi Normal

Dengan taraf signifikansi sebesar 5% , maka:

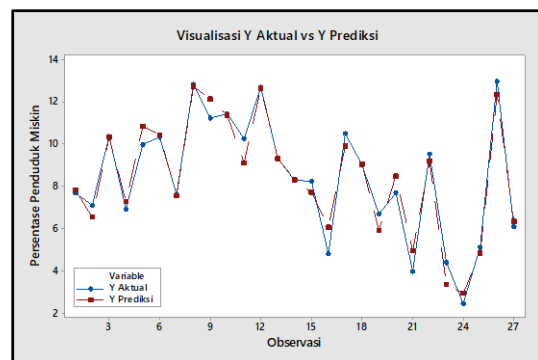


Gambar 3. Uji Normalitas Residual

Berdasarkan Gambar 3, secara visual dapat diketahui bahwa titik-titik persebaran residual memiliki kecenderungan tersebar di persekitaran garis dan cenderung mengikuti garis linear, hal ini didukung dengan nilai P -Value yang didapatkan lebih besar dari 0.05, sedemikian sehingga dapat disimpulkan jika asumsi dari residual berdistribusi normal telah terpenuhi.

3.6 Pembahasan

Model regresi nonparametrik dengan estimator *Spline Truncated* dapat digunakan untuk memprediksi suatu nilai dari variabel respon berdasarkan variabel prediktor. Adapun visualisasi antara y aktual dengan y prediksi dari model regresi nonparametrik *spline truncated* ditampilkan pada Gambar 4.



Gambar 4. Visualisasi y aktual dengan y prediksi

Berdasarkan gambar 4, diketahui bahwa hasil prediksi menggunakan model regresi nonparametrik dengan estimator *Spline Truncated* terlihat cenderung mengikuti pola data aktual, yang artinya model yang dibangun dapat dikatakan cukup baik untuk memprediksi persentase penduduk miskin Kabupaten/Kota di Provinsi Jawa Barat. Dengan didukung oleh nilai koefisien determinasi sebesar 95.33%, yang artinya model regresi nonparametrik yang dibangun pada

penelitian ini mampu menjelaskan pengaruh variabel prediktor terhadap variabel respon yaitu Persentase Penduduk Miskin sebesar 95.33%.

Selanjutnya berdasarkan pengujian hipotesis baik simultan dan parsial, diketahui jika variabel prediktor yang digunakan pada penelitian ini berpengaruh secara signifikan terhadap variabel respon.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis dan pembahasan, maka dapat dituliskan kesimpulan sebagai berikut:

1. Pada pemodelan persentase penduduk miskin Kabupaten/Kota di Provinsi Jawa Barat Tahun 2020 menggunakan model regresi nonparametrik, diketahui bahwa model regresi regresi nonparametrik *Spline Truncated* yang terbaik adalah model dengan 3 titik knot. Hal ini ditunjukkan oleh nilai GCV yang paling minimum yaitu sebesar 2.14, dan nilai koefisien determinasi (R^2) sebesar 95.33%.
2. Berdasarkan hasil pengujian hipotesis secara simultan, diperoleh nilai F_{hit} sebesar 34.44 dimana nilai tersebut lebih besar daripada nilai $F_{(0.05,16,10)}$ sebesar 2.83, sehingga diputuskan Tolak H_0 , yang artinya variabel prediktor secara simultan mempengaruhi variabel respon.
3. Berdasarkan hasil pengujian hipotesis secara parsial, dapat disimpulkan bahwa semua variabel berpengaruh terhadap variabel respon yaitu persentase penduduk miskin Kabupaten/Kota di Provinsi Jawa Barat.

Saran yang dapat diusulkan untuk penelitian selanjutnya adalah dapat mengembangkan model estimator campuran untuk setiap variabel prediktor berdasarkan karakteristik pola hubungannya dengan variabel respon.

DAFTAR PUSTAKA

- Budiantara, I. N. 2009. *Spline Dalam Regresi Nonparametrik dan Semiparametrik: Pidato Pengukuhan untuk Jabatan Guru Besar Institut Teknologi Sepuluh Nopember*. Surabaya.
- Bidari, D. R., dan Budiantara, I. N. 2020. Pemodelan Faktor yang Mempengaruhi Persentase Anak Putus Sekolah di Jawa Timur menggunakan Regresi Nonparametrik *Spline Truncated*. *Jurnal Sains dan Seni ITS*, 9(2), 115-122
- Budiantara, I. N. 2019. *Regresi Nonparametrik Spline Truncated*. ITS Press. Surabaya.
- Draper, N. dan Smith, H. 1992. *Applied Regression Analysis*. John Wiley & Sons, Inc. USA.

Eubank, R. L. 1999. *Nonparametric Regression and Spline Smoothing*. Marcel Dekker. New York.

Dani, A. T. R., Adrianingsih, N. Y., dan Ainurrochmah, A. 2020. *Pengujian Hipotesis Simultan Model Regresi Nonparametrik Spline Truncated dalam Pemodelan Kasus Ekonomi*. *JAMBURA Journal of Probability and Statistics*, 1(2), 98-106.

Dani, A. T. R., Adrianingsih, N. Y., Ainurrochmah, A., dan Sriningsih, R. 2021. *Flexibility of Nonparametric Regression Spline Truncated on Data without a Specific Pattern*. *Jurnal Litbang Edusaintech (JLE)*, 2(1), 37-43.

Dani, A. T. R., Ni'matuzzahroh, L., Ratnasari, V., dan Budiantara, I. N. 2021. *Pemodelan Regresi Nonparametrik Spline Truncated pada Data Longitudinal*. *INFERENSI*. 4(1), 47-55.

Dani, A. T. R. dan Adrianingsih, N. Y. 2021. *Pemodelan Regresi Nonparametrik dengan Estimator Spline Truncated dan Deret Fourier*. *JAMBURA Journal of Mathematics*, 3(1), 26-36.

Montgomery, D. C., Peck, E. A., dan Vining, G. G. 2012. *Introduction to Linear Regression Analysis: Fifth Edition*. John Wiley & Sons, Inc. USA.

Montoya, E., Ulloa, N., dan Miller, V. 2014. *A Simulation Study Comparing Knot Selection Methods with Equally Spaced Knots in a Penalized Regression Spline*. *International Journal of Statistics and Probability*, 03(03), 96-110.

Rahim, F., Budiantara, I. N., dan Permatasari, E. O. 2019. *Spline Truncated Nonparametric Regression Modeling for Maternal Mortality Rate in East Java*, *INFERENSI*. 2 (1), 39-44.

Saputro, D. R. S., Demu, K. R., dan Widyaningsih, P. 2018. *Nonparametric Truncated Spline Regression Model on Data of Human Development Index (HDI) in Indonesia*. *IOP Conf. Series: Journal of Physics*, 1188.

Sholikha, M., Susilawati, M., dan Srinadi, I. G. A. M. 2019. *Pemodelan Nilai Kurs terhadap Harga Saham pada Data Longitudinal Menggunakan Regresi Nonparametrik Spline Truncated*. *E-Jurnal Matematika*, 8(4), 259-263.

Yanthi, N. P. D., dan Budiantara, I. N. 2016. *Pemodelan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia menggunakan Regresi Nonparametrik Spline di Jawa Tengah*. *Jurnal Sains dan Seni ITS*, 5(2), 157-162.

Analisis Pengaruh Kepadatan Penduduk Dan Jumlah Lansia Terhadap Kasus Kematian Akibat Covid-19 di Provinsi DKI Jakarta Menggunakan *Multiple Classification Analysis* (MCA)

Habni Hamara Azmaty⁽¹⁾, Hussein Adi Permana⁽²⁾, Lisa Agustina⁽³⁾, Muhamad Fikri Ramdhani⁽⁴⁾, Naufal Abdul Rafi Zaqi⁽⁵⁾, Risni Julaeli Yuhan⁽⁶⁾

Politeknik Statistika STIS

Jl. Otto Iskandardinata Telp. (021) 8508812 Jakarta 13330

e-mail: 211810321@stis.ac.id⁽¹⁾

ABSTRAK

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) adalah virus yang menyebabkan penyakit Covid-19. Di Indonesia, jumlah kasus terkonfirmasi positif Covid-19 hingga 23 Mei 2021 adalah 1.8 juta orang dengan jumlah kasus kematian mencapai 49 ribu orang. DKI Jakarta merupakan salah satu provinsi dengan kasus kematian akibat Covid-19 tertinggi di Indonesia. Tujuan dari penelitian ini adalah untuk menganalisis pengaruh kepadatan penduduk dan jumlah penduduk lansia terhadap jumlah kasus kematian akibat Covid-19 di DKI Jakarta. Metode analisis yang digunakan adalah *Multiple Classification Analysis* (MCA) dengan unit observasi sebanyak 267 kelurahan yang berada di Provinsi DKI Jakarta. Hasil penelitian ini menunjukkan bahwa kepadatan penduduk dan jumlah penduduk lansia berpengaruh secara signifikan terhadap jumlah kasus kematian akibat Covid-19 di Provinsi DKI Jakarta. Semakin tinggi kepadatan penduduk maka jumlah kematian akibat Covid-19 cenderung meningkat. Begitu juga dengan jumlah penduduk lansia, semakin tinggi jumlah penduduk lansia, maka jumlah kasus kematian akibat Covid-19 juga cenderung meningkat.

Kata kunci : *Covid-19, Multiple Classification Analysis (MCA), Kepadatan Penduduk, Jumlah Penduduk Lansia.*

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a virus that causes the Covid-19 disease. In Indonesia, the number of positive confirmed cases of Covid-19 until May 23 2021 has reached 1.8 million people with the number of deaths reaching 49 thousand people. DKI Jakarta is one of the provinces with the highest Covid-19 deaths in Indonesia. The aim of this study was to analyze the impact of population density and the number of elderly residents on the number of cases of death due to Covid-19 in DKI Jakarta. The analytical method used is *Multiple Classification Analysis* (MCA) with 267 sub-districts of observation units in DKI Jakarta Province. The results of this study indicate that population density and the number of elderly residents have a significant effect on the number of deaths due to Covid-19 in DKI Jakarta Province. The higher the population density, the number of deaths due to Covid-19 tends to increase. Likewise with the number of elderly people, the increase in the number of elderly people, the number of cases of death due to Covid-19 also tends to increase.

Keywords : *Covid-19, Multiple Classification Analysis (MCA), Population Density, Number of Elderly Population.*

1. PENDAHULUAN

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) merupakan jenis baru dari virus corona yang menular dari manusia ke manusia dan menjadi penyebab dari munculnya penyakit yang disebut Covid-19 (WHO, 2020). Virus ini pertama kali diidentifikasi keberadaannya pada akhir bulan Desember 2019 di Kota Wuhan, China. Virus ini dapat menyebabkan beberapa penyakit pada manusia, seperti gangguan sistem pernafasan dan infeksi berat pada paru-paru, hingga

yang terparah yaitu kematian. Virus corona dapat menjangkiti siapa saja, tanpa memandang jenis kelamin ataupun usia seseorang. Berdasarkan data yang bersumber dari Gugus Tugas Percepatan Penanganan Covid-19 Republik Indonesia, jumlah kasus terkonfirmasi positif hingga 23 Mei 2021 adalah 1.8 juta orang dengan jumlah kasus kematian mencapai 49 ribu orang. Tingkat kematian yang diakibatkan oleh Covid-19 sebesar 2.78%. Covid-19 telah menyebar di seluruh provinsi di Indonesia. Provinsi yang tercatat memiliki angka

kasus positif Covid-19 tertinggi yaitu DKI Jakarta dengan jumlah 347 ribu orang hingga 23 Mei 2021. Selain itu, Provinsi DKI Jakarta menempati peringkat ketiga per 23 Mei 2021 dengan jumlah kasus kematian akibat Covid-19 terbanyak di Indonesia mencapai 6.3 ribu orang. Hal ini terjadi karena Provinsi DKI Jakarta merupakan episentrum atau pusat penularan Covid-19 di Indonesia. Mengingat Provinsi DKI Jakarta juga berperan sebagai pusat pemerintahan sekaligus pusat perekonomian di Indonesia, sehingga ada banyak orang yang masuk dan keluar DKI Jakarta.

Provinsi DKI Jakarta memiliki kepadatan penduduk sebesar 16.2 ribu jiwa/km². Angka tersebut menempatkan DKI Jakarta sebagai provinsi dengan kepadatan penduduk tertinggi di Indonesia. Menurut penelitian Bhadra (2021) tentang pengaruh kepadatan penduduk terhadap penyebaran Covid-19 dan kematian akibat Covid-19 di India, terdapat korelasi positif dan pengaruh signifikan antara kepadatan penduduk terhadap penyebaran Covid-19 dan berkaitan langsung dengan kematian akibat Covid-19. Artinya, semakin tinggi kepadatan penduduk suatu daerah, semakin tinggi pula penyebaran Covid-19 dan kematian akibat Covid-19. Daerah dengan kepadatan penduduk tinggi seperti DKI Jakarta, sangat dimungkinkan terjadi interaksi yang sangat masif antar penduduknya, terutama pada tempat umum seperti bandara, stasiun, sekolah, pasar, dan lain-lain. Semua tempat umum tersebut dapat menciptakan kerumunan sehingga akan memperbesar peluang penyebaran Covid-19 yang berkaitan langsung dengan kasus kematian akibat Covid-19.

Faktor lain yang dapat memengaruhi kasus kematian akibat Covid-19 yaitu jumlah penduduk lanjut usia. Penelitian dari Alshogran & dkk (2021) di 113 negara yang menggunakan model *adjusted regression* mengungkapkan bahwa tingkat kematian akibat Covid-19, salah satunya dipengaruhi oleh persentase penduduk yang berusia 60 tahun keatas. Hasil yang tidak jauh berbeda juga dikemukakan Sannigrahi & dkk (2020). Penelitiannya mengungkapkan beberapa temuan bahwa terdapat pengaruh yang signifikan antara komposisi demografi terhadap kasus Covid-19 dan kematian akibat Covid-19, serta penduduk usia lanjut merupakan kelompok penduduk yang paling terdampak buruk akibat Covid-19. Selain itu, penelitian Ioannidis (2020) yang mengkaji data dari 14 negara dan 13 negara bagian di Amerika Serikat mengungkapkan bahwa penduduk yang berusia diatas 65 tahun memiliki risiko untuk meninggal akibat Covid-19 yang lebih tinggi dibandingkan penduduk lainnya. Berdasarkan data Gugus Tugas

Percepatan Penanganan Covid-19 RI, pada tingkat nasional, kematian akibat Covid-19 juga lebih banyak terjadi pada kelompok penduduk yang berusia 65 tahun keatas dengan persentase sebesar 39.15 persen. Hal tersebut dapat disebabkan oleh faktor imunitas tubuh pada penduduk lanjut usia yang lebih rendah dibandingkan penduduk yang lebih muda. Dengan imunitas yang rendah, penduduk usia lanjut akan menerima dampak kesehatan yang lebih parah dibandingkan penduduk yang lebih muda sehingga dapat meningkatkan risiko kematian.

Hubungan kausalitas antarvariabel umumnya dianalisis menggunakan regresi linier berganda. Kesulitan yang biasa ditemui dalam analisis regresi linier berganda adalah ketika model tersebut tidak memenuhi asumsi klasik sebagai syarat model yang *Best Linier Unbiased Estimator* (BLUE). Model yang tidak memenuhi asumsi klasik kebanyakan adalah model yang mengandung *dummy* variabel. Untuk mengatasi hal tersebut, Frank M. Andrews pada tahun 1973 mengembangkan metode yang bernama *Multiple Classification Analysis* (MCA) (Sugiarto, 2018).

MCA merupakan metode yang lebih tepat digunakan untuk mengetahui hubungan kausalitas dengan kondisi variabel bebas berskala *dummy* atau kategori dan variabel tidak bebas memiliki skala ukur interval/rasio dibanding analisis regresi linier dengan *dummy* variabel. Pada analisis regresi dengan *dummy* variabel, salah satu kategori harus digunakan sebagai kategori referensi sehingga kategori tersebut dianggap tidak memiliki pengaruh terhadap model, sedangkan model MCA menyertakan semua kategori dari setiap variabel bebasnya.

Penelitian dari Sugiarto (2018) yang mengkaji tentang penggunaan metode MCA sebagai metode alternatif pada penelitian dengan variabel bebas berjenis kategorik menyimpulkan bahwa koefisien serta output yang berasal dari penggunaan metode MCA lebih mudah untuk diterapkan dan dipahami. Jika dibandingkan dengan regresi linier berganda, hasil dari metode MCA juga tidak jauh berbeda.

Berdasarkan penjelasan di atas, penelitian ini bertujuan untuk menganalisis pengaruh kepadatan penduduk dan jumlah penduduk lansia terhadap jumlah kasus kematian akibat Covid-19 di Provinsi DKI Jakarta dengan menggunakan metode *Multiple Classification Analysis* (MCA).

2. METODE PENELITIAN RUANG LINGKUP PENELITIAN

Variabel tidak bebas dalam penelitian ini adalah jumlah kasus kematian akibat Covid-19 hingga tanggal 23 Mei 2021. Sementara itu, variabel bebas yang digunakan adalah kepadatan penduduk dan jumlah penduduk lansia. Penelitian ini menggunakan metode *Multiple Classification Analysis* (MCA) dengan unit observasi sebanyak 267 kelurahan yang berada di Provinsi DKI Jakarta.

Tabel 1. Definisi Operasional Variabel

Variabel	Nama	Kategori	Kode
Tidak Bebas	Jumlah Kasus Kematian Akibat Covid-19	(satuan jumlah)	-
Bebas	Kepadatan Penduduk	Rendah	1
		Tinggi	2
	Jumlah Penduduk Lansia	Rendah	1
		Sedang	2
		Tinggi	3

Klasifikasi variabel kepadatan penduduk dibagi menjadi 2 kategori. Kepadatan penduduk dikategorikan rendah jika nilainya kurang dari median data. Kepadatan penduduk dikategorikan tinggi jika nilainya lebih dari sama dengan median data. Sementara itu, kategorisasi variabel jumlah penduduk lansia dilakukan dengan menggunakan nilai kuartil dari data. Kategorisasi tersebut adalah sebagai berikut (Hasanah, 2020):

Tabel 2. Kategorisasi Variabel Jumlah Lansia

No.	Interval Data	Kategori
1	$Data < Q_1$	Rendah
2	$Q_1 \leq Data \leq Q_3$	Sedang
3	$Data > Q_3$	Tinggi

SUMBER DATA

Data yang disertakan dalam penelitian ini berasal dari berbagai sumber yaitu:

Tabel 3. Sumber Data Variabel Penelitian

Variabel	Nama	Skala Pengukuran	Sumber
Tidak Bebas	Jumlah Kasus Kematian Akibat Covid-19	Rasio	Situs Riwayat File Covid-19 DKI Jakarta
Bebas	Kepadatan Penduduk	Ordinal	Situs Open Data Jakarta
	Jumlah Penduduk Lansia	Ordinal	Situs Open Data Jakarta

MULTIPLE CLASSIFICATION ANALYSIS (MCA)

MCA diperkenalkan oleh Yates pada tahun 1934, kemudian dilakukan pengembangan lebih lanjut pada tahun 1952 oleh Anderson Bancroft. Ruang lingkup dari metode ini lebih luas dibandingkan dengan Model Regresi Linier. Tujuan MCA adalah untuk melihat besaran pengaruh setiap variabel bebas terhadap variabel tak bebas, dengan mempertimbangkan atau tidak mempertimbangkan pengaruh variabel bebas lainnya. MCA digunakan pada kondisi variabel bebas dengan skala kategorik dan variabel tidak bebasnya memiliki skala ukur rasio/interval. Adapun keunggulan dari MCA adalah sebagai berikut:

1. Dapat menunjukkan pengaruh setiap variabel bebas terhadap variabel tak bebas baik sebelum maupun sesudah disesuaikan dengan variabel bebas lain.
2. Dalam persamaan MCA, pengaruh setiap variabel bebas merupakan nilai rata-rata dari variabel tidak bebas setelah disesuaikan dengan variabel bebas lainnya.
3. Koefisien MCA dihasilkan untuk setiap kategori pada variabel bebasnya, yaitu dalam bentuk variabel berskala kategorik dan dinyatakan sebagai *deviation* (penyimpangan) dari angka rata-ratanya.
4. MCA tidak memiliki batasan skala ukur untuk variabel bebasnya, artinya variabel bebas yang berskala ukur terendah (nominal) dapat digunakan.

Pada dasarnya, model MCA merupakan model penambahan atau model aditif yaitu sebagai berikut (Sugiarto, 2018) :

$$Y_{ij\dots n} = \bar{Y} + a_i + b_j + \dots + \varepsilon_{ij\dots n} \quad (1)$$

Dimana:

$Y_{ij\dots n}$: nilai amatan untuk individu berkategori ke-i pada variabel A, kategori ke-j pada variabel B, dan seterusnya

\bar{Y} : rata-rata total (*grandmean*) dari variabel tidak bebas

a_i : konstanta dari pengaruh kategori ke-i variabel A

b_j : konstanta dari pengaruh kategori ke-j variabel B

$\varepsilon_{ij\dots n}$: *error* dari individu yang selaras dengan $Y_{ij\dots n}$

Terdapat beberapa syarat yang harus dipenuhi dalam analisis MCA, diantaranya :

1. Data yang dijadikan bahan analisis merupakan data individu.
2. Tidak terdapat *outlier* dalam variabel tak bebas.
3. Tidak adanya interaksi yang terjadi antar variabel bebas.

Langkah-langkah dalam melakukan analisis MCA adalah sebagai berikut (Sugiarto, 2018):

1. Melakukan observasi/pengamatan pada data untuk melihat ada atau tidaknya *outlier* pada variabel tidak bebas.
2. Melakukan Uji Simultan dan Uji Parsial untuk mengetahui keberadaan pengaruh variabel bebas terhadap variabel tidak bebas. Pengujian ini dilakukan berdasarkan nilai *F main effect*.

Hipotesis yang disertakan pada pengujian ini adalah sebagai berikut:

H_0 = Tidak terdapat pengaruh variabel bebas terhadap variabel tak bebas

H_1 = Minimal terdapat satu variabel bebas yang berpengaruh terhadap variabel tidak bebas

Statistik uji pada pengujian ini merupakan uji F, yaitu:

- a. Uji Simultan :

$$F = \frac{E/(C-P)}{(T-E)/(N+P-C-1)} \quad (2)$$

Keterangan:

F = nilai F hitung

C = jumlah kategori

P = jumlah variabel tak bebas

T = jumlah kuadrat total

E = *explained sum of squares*

- b. Uji Parsial :

$$F_\alpha = \frac{U_\alpha/(C_\alpha-1)}{(T-U_\alpha)/(N-C_\alpha)} \quad (3)$$

Keterangan:

F_α =nilai F hitung variabel ke a

T = *total sum of squares*

C_α = banyak kategori variabel ke-a

U_α = jumlah kuadrat antar kategori variabel ke-a

Keputusan: jika $p\text{-value} < \alpha$ (tingkat signifikansi) atau $F \text{ hitung} > F_{0,10(V_1, V_2)}$, ambil keputusan tolak H_0

3. Selanjutnya melakukan tahapan pengecekan keberadaan interaksi antar variabel bebas, tujuannya untuk menguji terpenuhinya asumsi linier aditif. Hipotesis yang digunakan adalah sebagai berikut :

H_0 = Tidak terdapat interaksi antar variabel bebas

H_1 = Terdapat interaksi antar variabel bebas

Terdapat tiga langkah untuk menguji keberadaan dari pengaruh interaksi (Susel, 2011).

- a. Menguji F interaksi yang didapatkan dari pembagian antara rata rata kuadrat interaksi antar variabel bebas dan rata rata kuadrat dari residual. Tolak H_0 jika $p\text{-value} < \alpha$ atau $F > F_{(C-P);(N-C+P-1)}$. Jika $p\text{-value} < 10\%$, maka pengujian dilanjutkan ke tahap kedua. Jika $p\text{-value} \geq 10\%$, maka asumsi linier aditif terpenuhi.

- b. Menghitung rasio:

$$Rasio = \frac{\text{sum of square interaksi}}{\text{sum of square main effect}} \times 100\%$$

Jika rasio $\leq 10\%$, maka asumsi linier aditif terpenuhi. Jika rasio $> 10\%$, maka pengujian dilanjutkan ke tahap ketiga.

- c. Menghitung rasio:

$$Rasio = \frac{\text{sum of square interaksi}}{\text{sum of square total}} \times 100\%$$

Jika rasio $\leq 1\%$, maka asumsi linier aditif terpenuhi. Jika rasio $> 1\%$, maka analisis MCA tidak dapat dilakukan.

4. Langkah selanjutnya adalah melihat keeratan hubungan dari setiap variabel bebas terhadap variabel tidak bebas dengan menggunakan nilai eta (η) dan beta (β). Perhitungan Eta (η) dan Beta (β) adalah sebagai berikut.

- a. Nilai eta untuk variabel a

Nilai eta menunjukkan pengaruh setiap variabel bebas terhadap variabel tidak bebas sebelum disesuaikan dengan variabel bebas lainnya (sebelum *adjustment*).

$$\eta_\alpha = \sqrt{\frac{U_\alpha}{T}} \quad (5)$$

$$U_\alpha = \sum_{i=1}^{C_\alpha} N(\bar{Y}_{\alpha i} - \bar{Y})^2 \quad (6)$$

$$T = \sum_k N(Y_k - \bar{Y})^2 \quad (7)$$

Keterangan:

η_α = nilai eta variabel a

T = sum of square total

b. Nilai beta untuk variabel a

Nilai beta menunjukkan pengaruh setiap variabel bebas terhadap variabel tidak bebas setelah disesuaikan dengan variabel bebas lainnya (sesudah *adjustment*).

$$\beta_\alpha = \sqrt{\frac{D_\alpha}{T}} \quad (6)$$

$$D_\alpha = \sum_{i=1}^{C_\alpha} N(\bar{Y}_{ai(adj)} - \bar{Y})^2 \quad (7)$$

$$T = \sum_k N(Y_k - \bar{Y})^2 \quad (8)$$

Keterangan:

β_α = nilai beta untuk variabel a

D_α = sum of square adjusted antar kategori variabel a

$\bar{Y}_{ai(adj)}$ = rata-rata dari variabel tidak bebas pada kategori ke-i variabel a *adjusted*

5. Koefisien determinasi (R^2) berfungsi untuk mengerahui besaran pengaruh dari variabel bebas terhadap variabel tidak bebasnya. Koefisien determinasi diperoleh dari:

$$R^2 = E/T \quad (9)$$

$$R_{adj}^2 = 1 - (1 - R^2)A \quad (10)$$

Keterangan:

E = Explained sum squares

T = Total sum squares

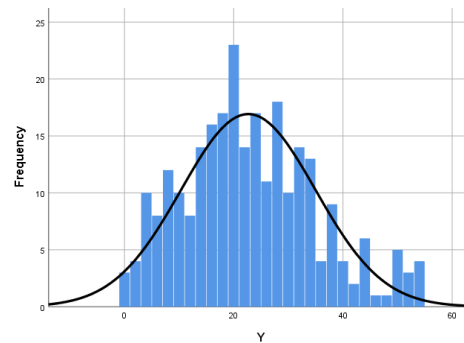
A = derajat bebas *adjusted*, dimana

$$A = \frac{N-1}{N+P-C-1} \quad (11)$$

3. HASIL DAN PEMBAHASAN

Sebelum melakukan analisis MCA, data yang diperoleh harus dilihat terlebih dahulu keberadaan *outlier*. Pendeketeksian *outlier* dilakukan dengan mencari nilai Z skor dari seluruh data pada variabel tak bebas. Menurut Santoso dalam Andriani (2018),

nilai Z skor yang berada di luar kurva normal -2,5 hingga 2,5 digolongkan menjadi data *outlier*. Jika didasarkan pada nilai Z skor dari 267 observasi, terdapat 6 observasi yang tergolong *outlier* sehingga peneliti mengambil keputusan untuk mengeluarkan 6 observasi tersebut. Maka, jumlah observasi yang digunakan untuk analisis MCA berjumlah 261 kelurahan. Dari hasil pengolahan *software* SPSS terhadap variabel tidak bebas yaitu jumlah kasus kematian akibat Covid-19 setiap kelurahan di Provinsi DKI Jakarta, didapatkan histogram sebagai berikut :



Gambar 1. Frekuensi jumlah kasus kematian Covid-19 di DKI Jakarta

Setelah melihat sebaran data, langkah selanjutnya adalah melakukan pengujian simultan dan parsial mengenai pengaruh variabel kepadatan penduduk dan jumlah penduduk lansia terhadap jumlah kasus kematian akibat Covid-19. Pengujian dilakukan dengan uji ANOVA. Dari hasil uji ANOVA, dapat dilihat model tersebut signifikan atau tidak secara simultan. Model yang signifikan menunjukkan bahwa terdapat paling sedikit satu variabel bebas yang secara signifikan memengaruhi jumlah kasus kematian akibat Covid-19. Selain itu, dari hasil uji ANOVA juga dapat dilihat variabel bebas mana yang signifikan memengaruhi jumlah kasus kematian akibat Covid-19 secara parsial. Hasil yang diperoleh adalah sebagai berikut :

Tabel 4. Hasil Pengujian Variabel Bebas dalam Model MCA

	SS	df	Mean Square	F	Sig.
(combined)	21789.243	3	7263.081	106.099	0.000
Kepadatan Penduduk	228.784	1	228.784	3.342	0.069
Jumlah Penduduk Lansia	19408.076	2	9704.038	141.757	0.000

Model	21789.243	3	7263.081	106.099	0.000
Residual	17593.087	257	68.456		
Total	39382.330	260	151.470		

Berdasarkan tabel. 4, *p-value* yang dihasilkan dari model secara simultan lebih kecil dibandingkan tingkat signifikansi 10% yaitu sebesar 0.000. Hal tersebut menunjukkan bahwa secara simultan terdapat paling sedikit satu variabel bebas yang berpengaruh terhadap jumlah kasus kematian akibat

Covid-19. Pengujian secara parsial terhadap seluruh variabel bebas menunjukkan bahwa dengan tingkat signifikansi 10%, semua variabel bebas, yaitu kepadatan penduduk dan jumlah penduduk lansia, memiliki pengaruh yang signifikan terhadap jumlah kasus kematian akibat Covid-19.

Tabel 5. Hasil Pengujian Interaksi Antar Variabel Bebas dalam Model MCA

	<i>SS</i>	<i>Df</i>	<i>Mean Square</i>	<i>Sig.</i>
<i>(combined)</i>	21789.243	3	7263.081	0.000
Kepadatan Penduduk * Jumlah Penduduk Lansia	391.026	2	195.513	0.057

Tahapan analisis selanjutnya yaitu menguji keberadaan pengaruh interaksi antar variabel bebas. Berdasarkan hasil uji interaksi pada tabel 5, didapatkan nilai *p – value* sebesar 0.057 sehingga dapat disimpulkan bahwa dengan tingkat signifikansi 10% terdapat interaksi di antara kedua variabel bebas tersebut. Hasil ini menunjukkan bahwa pengujian asumsi linier aditif harus dilanjutkan ke tahap 2, yaitu menggunakan perbandingan antara jumlah kuadrat interaksi antar variabel bebas dan jumlah kuadrat dari *main effect*.

Didapatkan nilai rasio sebesar 1.8%. Nilai tersebut kurang dari 10% sehingga interaksi antar variabel bebas dapat diabaikan atau asumsi linier aditif telah terpenuhi. Oleh karena itu, tidak perlu dilakukan uji interaksi tahap 3 yaitu menggunakan perbandingan antara jumlah kuadrat dari interaksi antar variabel bebas dan jumlah kuadrat total (Dewi & dkk, 2020).

$$Rasio = \frac{SS\ interaction}{SS\ combined} \times 100\% = 1.8\%$$

Tahapan selanjutnya yaitu melakukan analisis MCA untuk melihat pola hubungan antara variabel kepadatan penduduk dan jumlah penduduk lansia dengan jumlah kasus kematian akibat Covid-19. Hasil dari pengujian ditampilkan pada tabel berikut.

Tabel 6. Analisis MCA

		<i>Predicted Mean</i>		<i>Deviation</i>	
		<i>Unadjusted</i>	<i>Adjusted For Factors</i>	<i>Unadjusted</i>	<i>Adjusted For Factors</i>
Kepadatan Penduduk	Rendah	19.63	21.67	-3.032	-0.989
	Tinggi	25.67	23.64	3.009	0.982
Jumlah Penduduk Lansia	Rendah	9.79	10.31	-12.872	-12.353
	Sedang	23.15	22.94	0.488	0.276
	Tinggi	35.74	35.63	13.075	12.966

Hubungan variabel bebas dengan jumlah kasus kematian akibat Covid-19 di Provinsi DKI Jakarta dapat dilihat dari nilai *adjusted deviation*. Nilai *adjusted deviation* negatif menunjukkan bahwa kelurahan yang berada pada kategori tersebut memiliki jumlah kasus kematian akibat

Covid-19 yang lebih rendah dibanding rata-rata jumlah kasus kematian akibat Covid-19 di Provinsi DKI Jakarta.

Berdasarkan tabel 6, dapat diketahui bahwa kelurahan dengan kepadatan penduduk tinggi memiliki jumlah kasus kematian akibat Covid-19 yang lebih tinggi dibanding kelurahan dengan kepadatan penduduk rendah. Kelurahan dengan jumlah penduduk lansia tinggi akan memiliki jumlah kasus kematian akibat Covid-19 yang lebih tinggi dibanding kelurahan dengan jumlah penduduk lansia sedang dan rendah. Model MCA yang didapatkan adalah sebagai berikut:

$$Kasus\ Kematian_{ij} = 22.66 + \widehat{Kepadatan\ penduduk}_i + \widehat{Jumlah\ Penduduk\ Lansia}_i$$

Berdasarkan model MCA yang didapatkan, 22.66 adalah nilai rata-rata dari kasus kematian akibat Covid-19 di DKI Jakarta (*grandmean*). Koefisien model dari variabel kepadatan penduduk dan jumlah penduduk lansia diperoleh dari nilai *adjusted deviation*. Dari model tersebut dapat

dilakukan estimasi jumlah kasus kematian akibat Covid-19. Hasil dari estimasi dapat dilihat pada tabel berikut :

Tabel 7. Hasil Estimasi Model MCA

Kepadatan Penduduk	Jumlah Penduduk Lansia	\hat{Y}_{ij}
Rendah	Rendah	9.318
	Sedang	21.947
	Tinggi	34.637
Tinggi	Rendah	11.289
	Sedang	23.918
	Tinggi	36.609

Berdasarkan tabel 7, kelurahan yang memiliki jumlah kasus kematian akibat Covid-19 tertinggi adalah kelurahan dengan kepadatan penduduk dan jumlah penduduk lansia yang tinggi dengan rata-rata kematian 36 sampai 37 orang. Sementara itu, kelurahan yang memiliki jumlah kasus kematian akibat Covid-19 terendah adalah kelurahan dengan kepadatan penduduk dan jumlah penduduk lansia yang rendah dengan rata-rata kematian 9 sampai 10 orang.

Setelah dilakukan analisis model MCA, tahap berikutnya yaitu mencari variabel yang memiliki pengaruh tertinggi terhadap jumlah kasus kematian akibat Covid-19 di Provinsi DKI Jakarta. Hasil dari perhitungan disajikan pada tabel berikut :

Tabel 8. Nilai Eta, Beta, dan R^2

	Eta	Beta	R	R-Squared
		<i>Adjusted For Factors</i>		
Kepadatan Penduduk	0.246	0.080	0.744	0.553
Jumlah Penduduk Lansia	0.740	0.721		

Berdasarkan tabel 8, nilai eta dan beta terbesar dimiliki oleh variabel jumlah penduduk lansia. Hal tersebut menunjukkan bahwa variabel jumlah penduduk lansia merupakan variabel memiliki pengaruh terbesar terhadap jumlah kasus kematian akibat Covid-19 sebelum dan sesudah dilakukan penyesuaian dengan variabel bebas lainnya. Nilai eta dan beta yang dihasilkan dari variabel jumlah penduduk lansia tidak jauh berbeda, secara berturut-turut yaitu 0.740 dan 0.721. Hal tersebut

menunjukkan bahwa pengaruh dari variabel jumlah penduduk lansia terhadap jumlah kasus kematian akibat Covid-19 tidak jauh berbeda, baik sebelum maupun sesudah dilakukan penyesuaian dengan variabel bebas lainnya.

Hal yang berbeda ditunjukkan oleh variabel kepadatan penduduk, variabel tersebut memiliki nilai eta yang lebih besar dibandingkan nilai beta. Nilai eta dan beta dari variabel kepadatan penduduk secara berturut-turut yaitu 0.246 dan 0.080. Hal tersebut menunjukkan bahwa pengaruh dari variabel

kepadatan penduduk terhadap jumlah kasus kematian akibat Covid-19 sebelum dilakukannya penyesuaian dengan variabel bebas lain lebih besar dibandingkan setelah dilakukan penyesuaian dengan variabel bebas lainnya. Variabel kepadatan penduduk dan jumlah penduduk lansia, secara bersama-sama, dapat menjelaskan keragaman dari jumlah kasus kematian akibat Covid-19 sebesar 55.3% dan sisanya sebesar 44.7% dijelaskan oleh variabel lain yang tidak disertakan dalam model. Nilai *R-squared* yang tidak terlalu besar disebabkan oleh jenis data kategorik dari variabel bebas, sehingga belum mampu menjelaskan variabel tidak bebas yang berskala interval atau rasio (Sugiarto, 2018).

4. KESIMPULAN DAN SARAN

Berdasarkan berbagai analisis yang telah dilakukan, dapat disimpulkan bahwa kepadatan penduduk dan jumlah penduduk lansia berpengaruh signifikan terhadap jumlah kasus kematian akibat Covid-19 di Provinsi DKI Jakarta. Kelurahan dengan kepadatan penduduk tinggi memiliki jumlah kasus kematian akibat Covid-19 yang lebih tinggi dibanding kelurahan dengan kepadatan penduduk rendah. Kelurahan dengan jumlah penduduk lansia tinggi akan memiliki jumlah kasus kematian akibat Covid-19 yang lebih tinggi dibanding kelurahan dengan jumlah penduduk lansia sedang dan rendah. Jumlah penduduk lansia memiliki pengaruh yang lebih besar terhadap jumlah kasus kematian akibat Covid-19 di Provinsi DKI Jakarta dibanding kepadatan penduduk. Secara bersama-sama, kepadatan penduduk dan jumlah penduduk lansia dapat menjelaskan keragaman jumlah kasus kematian akibat Covid-19 sebesar 55,3%, sedangkan 44,7% lainnya dijelaskan oleh variabel lain yang tidak disertakan dalam penelitian ini.

DAFTAR PUSTAKA

- Alshogran, & dkk. 2021. *Predictors of Covid-19 case fatality rate: An ecological study*. Annals of Medicine and Surgery, 102319.
- Andriani, R. 2018. *Hubungan Kebiasaan Belajar Dengan Prestasi Belajar Siswa Sekolah Menengah Atas (SMA) Pertiwi 1 Padang*. TERAPUTIK: Jurnal Bimbingan dan Konseling, 2(1), 1-5.
- Bhadra, A., Mukherjee, A., & Sarkar, K. 2021. *Impact of population density on Covid-19 infected and mortality rate in India*. Modeling Earth Systems and Environment, 7(1), 623-629.

- Dewi, A. L., & dkk. 2020. *Analisis Negara-Negara Terjangkit Virus Covid-19 berdasarkan Kelompok Pendapatan dan Negara Tropis di Benua Asia Menggunakan Multiple Classification Analysis*. Jurnal Matematika Dan Statistika Serta Aplikasinya, Vol 8(1).
- Gugus Tugas Percepatan Penanganan COVID-19. <https://covid19.go.id/peta-sebaran-covid19>, diakses 23 Mei 2021
- Hasanah, M. 2020. *Pemilihan Jumlah Kategori Terbaik Pada Model Rough-Regresi Berdasarkan Mean Square Error (Studi Kasus: Tiga Variabel Bebas Numerik)* (Skripsi, Universitas Islam Negeri Sultan Syarif Kasim Riau).
- Ioannidis, J. P., Axfors, C., & Contopoulos-Ioannidis, D. G. 2020. *Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters*. Environmental research, 188, 109890.
- Sannigrahi, S., & dkk. 2020. *The overall mortality caused by covid-19 in the european region is highly associated with demographic composition: A spatial regression-based approach*. arXiv preprint arXiv:2005.04029.
- Sugiarto, S. 2018. *Multiple Classification Analysis (Mca) Sebagai Metode Alternatif Analisis Data Untuk Variabel Bebas Yang Kategori*. Jurnal Statistika Universitas Muhammadiyah Semarang, 6(2).
- Susel, A. 2011. *Multiple Classification Analysis: Theory and Application to Demography*. Journal of Acta Universitatis Lodziensis. 255, 183-189.
- World Health Organization. 2020. *Novel Coronavirus* <https://www.who.int/indonesia/news/novel-coronavirus/qa/qa-for-public>, diakses 24 Mei 2021

Membandingkan Seleksi Variabel Pada Data *Microarray* Menggunakan *Important Variable Value* dan *Genetic Algorithm* (Studi Kasus *Lung Cancer Dataset* dan *Prostate Cancer Dataset*)

Diana Nurlaily⁽¹⁾, Farida Nur Hayati⁽²⁾, Elly Pusporani⁽³⁾

Institut Teknologi Kalimantan

Jl, Soekarno-Hatta Km. 15 Telp./Fax. (0542) 8530801 Balikpapan 76127

e-mail: diana.nurlaily@lecturer.itk.ac.id, farida.nur@lecturer.itk.ac.id dan elly.pusporani@lecturer.itk.ac.id

ABSTRAK

Teknologi DNA microarray menarik minat yang luar biasa di kalangan komunitas ilmiah maupun kalangan industri. Ukuran sampel kecil dengan dimensi tinggi adalah tantangan utama untuk analisis data microarray. Banyak penelitian yang berkaitan dengan data microarray misalnya menyelidiki mekanisme genetik kanker, mengklasifikasikan jenis kanker atau membedakan antara jaringan kanker dan non-kanker. Penelitian tersebut bertujuan untuk menghasilkan kesimpulan dan interpretasi yang bermanfaat dari kumpulan data yang kompleks. Dalam penelitian ini, data yang digunakan adalah data *lung cancer* sebanyak 24257 variabel dan data *prostate cancer* sebanyak 12626 variabel. Data tersebut dianalisis dengan dua metode seleksi variabel yaitu *important variable value* dan *genetic algorithm* untuk meningkatkan akurasi klasifikasi data. *Important variable value* merupakan metode seleksi variabel dimana variabel dipilih berdasarkan tingkat kepentingan variabel terhadap model klasifikasi. Sedangkan metode *genetic algorithm* memilih variabel berdasarkan *fitness value*, kombinasi dan *cross over* sehingga menghasilkan kombinasi kumpulan variabel baru. Berdasarkan hasil analisis seleksi variabel pada data *lung cancer*, didapatkan jumlah variabel terpilih sebanyak 112 variabel dengan metode *important variable value*. Sedangkan metode *genetic algorithm* didapatkan jumlah variabel terpilihnya sebanyak 12266 variabel. Pada data *prostate cancer*, didapatkan jumlah variabel terpilih sebanyak 299 variabel dengan metode *important variable important*. Sedangkan metode *genetic algorithm* didapatkan jumlah variabel terpilihnya sebanyak 6359 variabel.

Kata kunci : *Microarray, important variable value, genetic algorithm, seleksi variabel*

ABSTRACT

DNA microarray technology is attracting tremendous interest in both the scientific community and industry. Small sample size with high dimensions is a significant challenge for microarray data analysis. Many studies are related to microarray data, for example, investigating the genetic mechanism of cancer, classifying cancer types, or distinguishing between cancerous and non-cancerous tissues. The study aims to generate valuable conclusions and interpretations from complex data sets. In this study, the data used were lung cancer data as many as 24257 variables and prostate cancer data as many as 12626 variables. The data were analyzed using two variable selection methods, namely important variable value and genetic algorithm, to increase the accuracy of data classification. Important variable value is a variable selection method where the variable is selected based on the level of importance of the variable to the classification model. In comparison, the genetic algorithm method selects variables based on fitness values, combinations, and cross-overs to produce new combinations of variables. Based on the variable selection analysis results on lung cancer data, the number of selected variables was 112 variables using the important variable value method. In comparison, the genetic algorithm method obtained the number of selected variables as many as 12266 variables. In prostate cancer data, the number of selected variables is 299 variables using the important variable important method. In contrast, the genetic algorithm method obtained the number of selected variables as many as 6359 variables.

Keywords : *Feature selection, microarray, important variable value, genetic algorithm*

1. PENDAHULUAN

Teknologi DNA *microarray* menarik minat yang luar biasa baik di kalangan komunitas ilmiah maupun kalangan industri. Kemampuan mengukur secara bersamaan aktivitas dan interaksi ribuan gen, menjadikannya sebagai pengetahuan baru mengenai mekanisme sistem kehidupan. Data *microarray* sendiri mempunyai ciri-ciri yaitu sampel kecil, dimensi tinggi, noise tinggi, redundansi tinggi dan distribusi kelas tidak seimbang (Wang & Simon, 2011). Apabila data tersebut dianalisis dengan metode yang tepat, maka akan dihasilkan pengetahuan baru yang bermanfaat dibidang tertentu misalnya biologi dan kedokteran. Banyak penelitian yang telah dirancang berkaitan dengan data *microarray* misalnya untuk menyelidiki mekanisme genetik kanker, dan untuk mengklasifikasikan berbagai jenis kanker atau membedakan antara jaringan kanker dan non-kanker. Semua penelitian ini bertujuan untuk menghasilkan kesimpulan dan interpretasi yang bermanfaat dari kumpulan data yang kompleks (Hira & Gillies, 2015).

Ukuran sampel kecil dengan dimensi tinggi adalah tantangan utama analisis menggunakan data *microarray*. Dimensi data *microarray* yang dapat mencapai 450.000 variabel membuat analisis menjadi kurang efektif untuk diproses dalam program komputer. Berdasarkan permasalahan tersebut maka perlu dilakukan analisis dengan mengurangi dimensi/variabel. Hal yang biasanya dilakukan untuk mengurangi dimensi suatu data *microarray* adalah menggunakan seleksi variabel. Seleksi variabel bekerja dengan cara menghilangkan variabel yang berlebih dan kurang relevan. Beberapa metode seleksi variabel yang dapat digunakan antara lain adalah *important value*, *Filter method* dan *Genetic algorithm*.

Penelitian tentang seleksi variabel data *microarray* telah banyak dilakukan antara lain oleh oleh Hambali, Oladele, & Adewole (2020) tentang seleksi variabel pada data kanker *microarray*, menghasilkan kesimpulan bahwa teknik seleksi variabel *hybrid* dapat meningkatkan performa akurasi yang lebih baik dari model klasifikasi penyakit kanker. Selain itu Rasmita Dash (2020) meneliti seleksi variabel dan klasifikasi data *microarray* berdasarkan ranking variabel.

Dalam penelitian ini, data yang digunakan adalah data kanker paru-paru sebanyak 24257 Variabel dan data kanker prostat sebanyak 12626 Variabel. Data tersebut kemudian akan dianalisis dengan dua metode seleksi variabel yaitu *important variable value* dan *genetic algorithm*. untuk

memilih dimensi atau variabel data sehingga dapat meningkatkan akurasi klasifikasi data. Metode *important variable value* termasuk dalam kategori *embedded* seleksi variabel dan *genetic algorithm* termasuk *wrapper* seleksi variabel. Pada tahun 2021 Wang, dkk melakukan penelitian terkait seleksi variabel menggunakan *important variable value*, yaitu memilih variabel berdasarkan kepentingannya. *Important variable value* merupakan salah satu fitur seleksi variabel yang ada di metode klasifikasi menggunakan *decision tree* atau *random forest*. Sedangkan Sayed dkk pada tahun 2019 meneliti seleksi variabel menggunakan *genetic algorithm*, dimana didapatkan seleksi variabel menggunakan KNN dan *random forest*.

Sehingga tujuan dari penelitian ini adalah Mengkaji metode seleksi variabel untuk mengklasifikasikan data *microarray*, dan Membandingkan metode *important variable value* dan *genetic algorithm* untuk seleksi variabel data *microarray*.

2. METODE PENELITIAN

2.1 Important Variable Value

Important variable value merupakan salah satu metode untuk memilih variabel. Metode *important variable* memilih variabel berdasarkan tingkat kepentingannya. Pada tahun 1984 Brieman, dkk mengusulkan *important variable* dengan menggunakan pemisahan untuk mengurangi resiko bahwa variabel yang penting disamarkan. *Important variable* mewakili signifikasin dari setiap variabel dalam data pengaruhnya terhadap model yang dihasilkan. Setiap variabel akan diberi peringkat sesuai dengan kontribusinya pada model. Sehingga dengan metode ini dapat menghilangkan variabel tertetu yang tidak memberikan kontribusi terhadap model. Nilai dari *important variable* dihitung dari penjumlahan penurunan kesalahan saat setiap variabel dipisahkan. Nilai *important variabel* dibatasi antara 0 sampai 1.

Salah satu cara yang paling sering digunakan untuk dalam menentukan tingkat kepentingan variabel yaitu menggunakan *permutation importance*. Pengukuran *Important variable* dilakukan pada metode klasifikasi menggunakan *Random Forest*, dimana pada *Random Forest* memperkirakan kepentingan covariat dengan cara permutasi setiap variabel dan mengklasifikasikan ulang berdasarkan variabel yang telah dipermutasi (Rad, Koohkan, Fanaei, & Rad, 2015). Setiap Variabel diukur tingkat kepentingannya dengan cara semua variabel dipermutasi dan ukuran tingkat kepentingannya sebagai perbedaan dalam ukurasi dari prediksi yang disebabkan oleh permutasi variabel tersebut (Hjerpe, 2016).

2.2 Genetic Algorithm

Genetic Algorithm (GA) adalah metode yang digunakan untuk menemukan solusi pada permasalahan optimasi dan pencarian secara komputasi. Metode GA terinspirasi dari evolusi biologi yang meliputi pewarisan, mutasi, seleksi, dan cross over. Metode GA memungkinkan untuk menemukan solusi dari masalah yang tidak dapat ditangani oleh metode optimasi lain yang disebabkan oleh kurangnya kontinuitas, turunan, linieritas dan variable lainnya. Metode GA dapat menyelesaikan permasalahan data yang mempunyai dimensi tinggi dalam menyeleksi variable. Variabel diseleksi menggunakan nilai *fitness* dan kemudian dikombinasikan dengan *cross over* dan mutasi sehingga menghasilkan generasi baru. Genetic algorithm untuk seleksi variabel yang berbasis wrapper memberikan fungsi output dengan kemampuan generalisasi yang baik (Djellali & Adda, 2017).

Gen adalah istilah untuk menunjukkan variabel, alel adalah nilai gen atau variabel, yang bisa 1 atau 0. Ini adalah 1 yang berarti bahwa variabel dipilih secara acak dan memiliki nilai 0 yang berarti bahwa variabel tidak dipilih. Kromosom adalah pengumpulan gen dan individu yang merupakan salah satu solusi yang mungkin untuk pemilihan variabel (Nurlaily, et al., 2019). Prosedur seleksi variabel menggunakan GA adalah sebagai berikut:

1. Inisialisasi menciptakan dan menginisialisasi individu dalam populasi secara acak.
2. Fitness Assignment adalah mengevaluasi solusi yang dihasilkan untuk dipilih atau tidak. Jika kesalahan tinggi, kebugaran rendah. Individu dengan kebugaran tinggi memiliki peluang lebih tinggi untuk dipilih.
3. Seleksi adalah memilih individu yang akan bergabung kembali dengan generasi berikutnya. Operator pilihan memilih individu berdasarkan nilai kebugaran mereka.
4. Crossover menggabungkan kembali individu yang dipilih untuk menghasilkan populasi baru. Operator ini secara acak memilih dua individu dan menggabungkan variabel mereka untuk mendapatkan keturunan dalam populasi baru.
5. Mutasi adalah solusi untuk masalah crossover yang menghasilkan keturunan yang sama. Dalam proses perubahan mutasi nilai beberapa variabel secara acak.

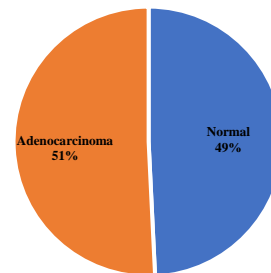
2.3 Data Microarray

Data microarray adalah data yang berisi serangkaian sampel bisa DNA, RNA, jaringan dan protein, sehingga jenis data microarray tergantung dari sampel yang diamati (Tuimala & Laine, 2003). Data microarray bisa memuat ratusan ribu gen yang

bisa diamati. Data microarray yang biasa digunakan dalam penelitian adalah DNA microarray. DNA microarray digunakan untuk menentukan tingkat ekspresi gen dalam sampel dan urutan gen dalam sampel. Data microarray yang digunakan pada penelitian ini ada dua yaitu data lung cancer dan prostate cancer. Perlu dilakukan serangkaian penelitian untuk mendapatkan data microarray dan disebut dengan microarray experiment.

3. HASIL DAN PEMBAHASAN

Data *lung cancer* pada penelitian ini terdiri dari 65 pengamatan pasien kanker paru-paru dengan jumlah ekspresi gen yang diamati sebanyak 24.257. Pada data lung cancer, pasien dibedakan menjadi dua yaitu pasien normal dan pasien *adenocarcinoma*. Karakteristik data *lung cancer* berdasarkan kelasnya dapat dilihat di Gambar 1.



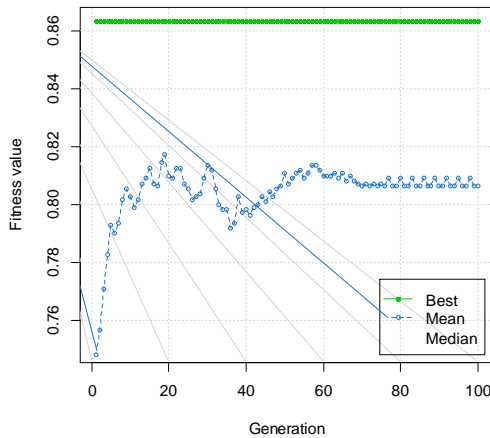
Gambar 1. Karakteristik data lung cancer berdasarkan kelas pasien

Berdasarkan Gambar 5.1 terlihat bahwa dari 65 pasien *lung cancer* terdapat 51% pasien masuk kelas normal dan 51% pasien termasuk kelas *adenocarcinoma*. Sehingga dilihat dari persebaran kelas pasien dapat dikatakan bahwa data *lung cancer* ini *balance* atau seimbang untuk setiap kelasnya.

Pada metode *important value* ini, menggunakan metode *decision tree* untuk metode klasifikasinya. Pada proses klasifikasi menggunakan *decision tree* ini terdapat fitur untuk seleksi variabel yaitu *important variable*, dimana saat proses klasifikasi, variabel yang diperkirakan mempengaruhi kelas *lung cancer* akan diseleksi berdasarkan tingkat kepentingannya. Seleksi variabel pada *lung cancer* dengan menggunakan metode *important value* didapatkan jumlah variabel yang terpilih sebanyak 112 variabel. Sehingga dengan menggunakan metode *important value* dapat mereduksi sebesar 99,53% dari total variabel keseluruhan. Metode *important value* ini memilih variabel berdasarkan nilai tingkat kepentingannya. Berdasarkan perhitungan didapatkan 112 variabel yang nilai tingkat kepentingannya tidak sama dengan 0

sehingga, variabel yang terpilih yaitu 112 variabel. Hal ini dikarenakan variabel yang nilai kepentingannya 0 dianggap tidak penting dan tidak perlu dimasukkan dalam model klasifikasi.

Seleksi variabel pada data *lung cancer* menggunakan metode *genetic algorithm* dimana iterasi yang dilakukan sebanyak 100 kali, menggunakan peluang *crossover* sebesar 0,8, peluang mutasi 0.05 dan *fitness value* yang digunakan adalah nilai kebaikan model AUC dari metode klasifikasi SVM. Berdasarkan iterasi yang dilakukan sebanyak 100 kali didapatkan nilai *fitness value* sebesar 0.863. berikut adalah nilai *fitness* yang didapatkan selama iterasi berlangsung.



Gambar 2. Nilai fitness GA data lung cancer

Menggunakan metode *genetic algorithm* didapatkan jumlah variabel terpilihnya sebanyak 12266 variabel. Jika dibandingkan dengan total variabel aslinya, metode *genetic algorithm* dapat mereduksi variabel sebanyak 49,43% dari jumlah variabel aslinya.

Sehingga jika dibandingkan hasil seleksi variabel untuk data *lung cancer* berdasarkan jumlah variabel terpilihnya dapat dilihat pada Tabel 1.

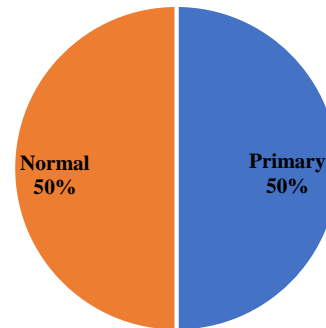
Tabel 1. Perbandingan seleksi variabel data lung cancer

data	Metode	Total Variabel	Variabel Terpilih	Persentase Variabel Terpilih
Lung Cancer	Important Value	24257	112	0,46%
	Genetic Algorithm		12266	50,57%

Pada setiap gambar harus diberikan keterangan di bawah gambar. Keterangan pada tabel diberikan di atas tabel. Keterangan dituliskan dengan huruf kecil kecuali pada karakter pertama pada tiap

kalimat. Seluruh gambar harus diberi penomoran secara berurutan. Jika Gambar besar maka diletakkan di tengah halaman (*center alignment*) dengan judul ditengah dan jika gambar kecil maka letakkan di tengah (*center columns*) baik itu pada kolom 1 ataupun pada kolom 2 dengan nama gambar rata *justify*, demikian halnya dengan tabel.

Data *prostate cancer* pada penelitian ini terdiri dari 124 pengamatan pasien kanker prostate dengan jumlah ekpresi gen yang diamati sebanyak 12.626. Pada data *prostate cancer*, pasien dibedakan menjadi dua yaitu pasien normal dan pasien *primary* (utama) . Karakteristik data *prostate cancer* berdasarkan kelasnya dapat dilihat di Gambar 3.

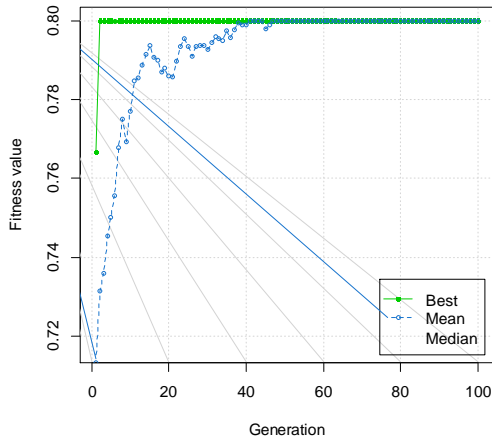


Gambar 3. Karakteristik data prostate cancer berdasarkan kelas pasien

Berdasarkan Gambar 3 terlihat bahwa dari 124 pasien *prostate cancer* terdapat 50% pasien masuk kelas normal dan 50% pasien termasuk kelas *primary*. Sehingga dilihat dari persebaran kelas pasien dapat dikatakan bahwa data *prostate cancer* ini *balance* atau seimbang untuk setiap kelasnya.

Pada metode *important value* ini, menggunakan metode decision tree untuk metode klasifikasinya. Pada proses klasifikasi menggunakan decision tree ini terdapat fitur untuk seleksi variabel yaitu *important variable*, dimana saat proses klasifikasi, variabel yang diperkirakan mempengaruhi kelas prostate cancer akan diseleksi berdasarkan tingkat kepentingannya. Sehingga diakhir proses seleksi variabel pada prostate cancer dengan menggunakan metode *important value* didapatkan jumlah variabel yang terpilih sebanyak 299 variabel. Metode *important value* ini memilih variabel berdasarkan nilai tingkat kepentingannya. Berdasarkan perhitungan didapatkan 299 variabel yang nilai tingkat kepentingannya tidak sama dengan 0 sehingga, variabel yang terpilih yaitu 299 variabel. Hal ini dikarenakan variabel yang nilai kepentingannya 0 dianggap tidak penting dan tidak perlu dimasukkan dalam model klasifikasi.

Seleksi variabel pada data *prostate cancer* menggunakan metode *genetic algorithm* dimana iterasi yang dilakukan sebanyak 100 kali, menggunakan peluang *crossover* sebesar 0.8, peluang mutasi 0.05 dan *fitness value* yang digunakan adalah nilai kebaikan model AUC dari metode klasifikasi SVM. Berdasarkan iterasi yang dilakukan sebanyak 100 kali didapatkan nilai *fitness value* yang paling optimum sebesar 0.800. berikut adalah nilai *fitness* yang didapatkan selama iterasi berlangsung.



Gambar 4. Nilai fitness GA data prostate cancer

Menggunakan metode *genetic algorithm* didapatkan jumlah variabel terpilihnya sebanyak 6359 variabel. Jika dibandingkan dengan total variabel aslinya, metode *genetic algorithm* dapat mereduksi variabel sebanyak 49,63% dari jumlah variabel aslinya.

Sehingga jika dibandingkan hasil seleksi variabel untuk data prostate cancer berdasarkan jumlah variabel terpilihnya dapat dilihat pada Tabel 2.

Tabel 2. Perbandingan seleksi variabel data prostate cancer

data	Metode	Total Variabel	Variabel Terpilih	Persentase Variabel Terpilih
Prostate Cancer	Important Value	12626	299	23,68%
	Genetic Algorithm		6359	50,37%

4. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis didapatkan kesimpulan bahwa data *lung cancer* dan *prostate cancer* merupakan data yang balance jika dilihat dari proporsi kelas pasien. Seleksi variabel pada data

lung cancer menggunakan metode *important value* memberikan hasil 122 variabel yang terpilih dan jika menggunakan metode *genetic algorithm* variabel terpilihnya sebesar 12.266 variabel. Seleksi variabel untuk data *prostate cancer* menggunakan metode *important value* jumlah variabel yang terpilihnya sebesar 299 variabel dan jika menggunakan metode *genetic algorithm* jumlah variabel terpilihnya sebanyak 6.359 variabel. Sehingga jika dilihat dari jumlah variabel terpilihnya metode *important value* memberikan reduksi variabel yang lebih tinggi dibandingkan metode *genetic algorithm* untuk data *lung cancer* dan *prostate cancer*.

Saran untuk penelitian ini adalah perlu dilakukan penelitian lanjutan untuk membuktikan metode seleksi variabel mana yang lebih baik jika dilihat dari nilai kebaikan model untuk mengklasifikasikan pasien dari data *lung cancer* dan *prostate cancer*. Metode yang bisa digunakan diantaranya adalah regresi logistik, *decision tree*, *SCM*, *Naïve Naves*.

DAFTAR PUSTAKA

Dash, R. (2020). A two stage grading approach for feature selection and classification of microarray data using Pareto based feature ranking techniques: A case study. *Journal of King Saud University –Computer and Information Sciences*, 232-247.

Djellali, C., & Adda, M. (2017). A New Predictive Approach to Variables Selection Through Genetic Algorith and Fuzzy Adaptive Resonance Theory Using Medical Diagnosis as a Case. *Procedia Computer Science*, 448-457.

Hambali, M. A., Oladele, T. O., & Adewole, K. S. (2020). Microarray cancer feature selection: Review, challenges and research directions. *International Journal of Cognitive Computing in Engineering* , 78-97.

Hjerpe, A., 2016. *Computing Random Forest Variable Importance Measures (VIM) on Mixed Continous and Categorical Data*. Stockholm, Sweden: KTH Royal Institute of Technology School of Computer Science and Communication.

Hira, Z. M., & Gillies, D. F. (2015). A Review of Feature Selection and Feature Extraction Methods. *Advaces in Bioinformatics* , 1-13.

Nurlailly, D., Irhamah, Purnami, S. W. & Kuswanto, H., 2019. *Support Vector Machine for Imbalanced Microarray Dataset Classification*

- Using Ant Colony Optimization and Genetic Algorithm*. s.l., AIP Publishing.
- Rad, M. R., Koohkan, S., Fanaei, H. R., & Rad, M. R. (2015). Application of Artificial Neural Networks to predict the final fruit weight and random forest to select important variables in native population of melon (*Cucumis melo* L.). *Scientia Horticulture*, 108-112.
- Sayed, S., Nassef, M., Badr, A., & farag, I. (2019). A Nested Genetic Algorithm for Feature Selection in High-dimensional Cancer Microarray Datasets. *Microarray Datasets*.
- Wang, L., Huang, Z., & Wang, R. (2021). Discrimination of cracked soybean seeds by near-infrared spectroscopy and random forest variable selection. *Infrared Physics and Technology*.
- Wang, X., & Simon, R. (2011). Microarray Based Cancer Prediction using Single Genes. *BMC Bioinformatics* , Vol. 12, hal. 391-400.

Pemodelan Regresi Spatial Autoregressive Fixed Effect Model Data Panel Pada Tingkat Kemiskinan Di Provinsi Jawa BaratMuhammad Irfan R⁽¹⁾, Teguh Ammar T⁽²⁾

Universitas Padjadjaran

Gedung Departemen Statistika, Jatinangor 45361, Sumedang, Jawa Barat

e-mail: muhammad18011@mail.unpad.ac.id

ABSTRAK

Kemiskinan menjadi salah satu isu dalam permasalahan global yang terjadi di semua negara berkembang termasuk negara Indonesia. Mengentaskan permasalahan kemiskinan tentunya menjadi hal yang utama karena sesuai dengan tujuan dalam mencapai pembangunan berkelanjutan atau *Sustainable Development Goals* (SDGs), karena pengentasan kemiskinan menjadi tujuan poin pertama yang diharapkan bisa tercapai. Kemiskinan juga menjadi salah satu permasalahan yang menjadi isu salah satu isu strategis RPJMD tahun 2018-2023 yang menjadi pusat perhatian pemerintah khususnya di Provinsi Jawa Barat yang merupakan provinsi dengan jumlah penduduk terbanyak di Indonesia. Penelitian ini dilakukan dalam menganalisis faktor-faktor yang mempengaruhi tingkat kemiskinan di Provinsi Jawa Barat. Tentunya tingkat kemiskinan tiap-tiap kabupaten/kota mempunyai tingkat yang berbeda-beda, maka terdapat kemungkinan adanya efek spasial dalam data. Maka dalam pemodelan tentunya harus mengakomodasi aspek spasial kemudian terbatasnya variabel yang dilibatkan karena keterbatasan data tentunya menimbulkan omitted variabel atau variabel yang relevan namun tidak ada dalam model maka digunakan pendekatan *fixed effect* model dalam mengatasi masalah tersebut. Sehingga pemodelan yang digunakan adalah *Spatial Autoregressive Fixed Effect* model (SAR-FEM). Hasil penelitian ini didapatkan yaitu Variabel Tingkat Pengangguran Terbuka, Indeks pembangunan Manusia dan Persentase Penduduk berpengaruh signifikan terhadap Tingginya tingkat kemiskina di Provinsi Jawa Barat. Model *spatial lag fixed effect* yang terbentuk dapat menjelaskan besarnya keragaman dari Tingkat Kemiskinan yang dapat dijelaskan oleh variabel prediktor sebesar 98.88% sedangkan 1.116% sisanya dijelaskan oleh variabel yang tidak dilibatkan kedalam pemodelan.

Kata kunci : *Kemiskinan, omitted variabel, Spatial Autoregressive Fixed Effect Model*

ABSTRACT

Poverty is one of the issues in global problems that occur in all developing countries, including Indonesia. Alleviating the problem of poverty is of course the main thing because it is in accordance with the objectives of achieving sustainable development or the Sustainable Development Goals (SDGs), because poverty alleviation is the first point goal that is expected to be achieved. Poverty is also one of the problems that is one of the strategic issues of the 2018-2023 RPJMD which is the center of government attention, especially in West Java Province which is the province with the largest population in Indonesia. This research was conducted to analyze the factors that influence the poverty level in West Java Province. Of course, the poverty level of each district/city has a different level, so there is a possibility that there is a spatial effect in the data. So in modeling, of course, having to accommodate spatial aspects and then the limited variables involved because of limited data, of course, causing omitted variables or relevant variables but not in the model, so a fixed effect model approach is used to overcome this problem. So that the modeling used is the *Spatial Autoregressive Fixed Effect* model (SAR-FEM). The results of this study showed that the Variable Open Unemployment Rate, Human Development Index and Population Percentage had a significant effect on the high level of poverty in West Java Province. The spatial lag fixed effect model that is formed can explain the magnitude of the diversity of the Poverty Level which can be explained by the predictor variable of 98.88% while the remaining 1.116% is explained by variables that are not included in the modeling.

Keywords : *Poverty, Omitted Variabel, Spatial Autoregressive Fixed Effect Model*

1. PENDAHULUAN

Perencanaan dalam sebuah pembangunan nasional merupakan sebuah langkah penting dalam mengantisipasi ketidakseimbangan yang terjadi dan memiliki sifat akumulatif. Artinya, perubahan yang terjadi akan menimbulkan perubahan pada sistem sosial. Perencanaan juga memiliki peranan yang sangat penting dalam sebuah proses pembangunan, salah satu peran perencanaan salsahturnya sebagai arahan dan proses pembangunan agar dapat berjalan mencapai sebuah tujuan yang ingin dicapai disamping sebagai indikator dalam mengukur kesuksesan proses pembangunan yang telah dilakukan. Sedangkan arti dari pembangunan itu sendiri yaitu upaya dalam mempercepat pertumbuhan ekonomi. (Harlik, 2013).

Kemiskinan merupakan sebuah keadaan dimana seseorang atau sekelompok orang, laki-laki dan juga perempuan yang tidak mampu memenuhi hak dasarnya seperti hak untuk mempertahankan dan mengembangkan hidup yang bermartabat. Hak-hak dasar masyarakat diantaranya adalah terpenuhi semua kebutuhan pangan, kesehatan, pendidikan, pekerjaan, perumahan, air bersih, sumber daya alam dan lingkungan hidup, rasa mendapatkan perlindungan, dan hak dalam bersosial politik.

Hal ini menjadi perhatian berbagai pihak untuk mengentaskannya. Perserikatan Bangsa-Bangsa (PBB) menjadikan “mengakhiri kemiskinan di manapun dan dalam bentuk apapun” sebagai tujuan pertama dalam Sustainable Development Goals (SDG’s) (BPS, 2016). Hal ini menunjukkan bahwa kemiskinan merupakan permasalahan utama yang harus segera diselesaikan oleh negara-negara di dunia. Demikian pula dengan pemerintah Indonesia yang menjadikan penanggulangan kemiskinan sebagai prioritas pembangunan nasional (BPS, 2020). Sedangkan pemerintah Provinsi Jawa Barat menjadikan kemiskinan sebagai permasalahan utama dalam lima tahun terakhir dan merupakan isu strategis pembangunan dalam lima tahun ke depan (RPJMD, 2018-2023). Isu strategis dalam sebuah pembangunan nasional merupakan salah satu hal yang penting dan harus diperhatikan atau dikedepankan karena dampaknya signifikan bagi daerah, dalam pencapaian jangka menengah/panjang, dan bisa menentukan sebuah pencapaian dalam tujuan penyelenggaraan pemerintah daerah dimasa mendatang.

Dalam hukum Tobler, segala sesuatu akan saling memiliki hubungan satu dengan yang lain, namun sesuatu yang memiliki kedekatan khusus akan lebih berpengaruh dari pada yang jauh (Anselin, 1988). Data kemiskinan di Jawa Barat diambil dari tiap-tiap kabupaten/kota, maka dimungkinkan bahwa kemiskinan di kabupaten/kota satu akan dapat mempengaruhi kemiskinan di kabupaten/kota lain yang saling berdekatan.

Muhammad Irfan R¹, Teguh Ammar T²/
J Statistika Vol. 14, No. 1, (2021)

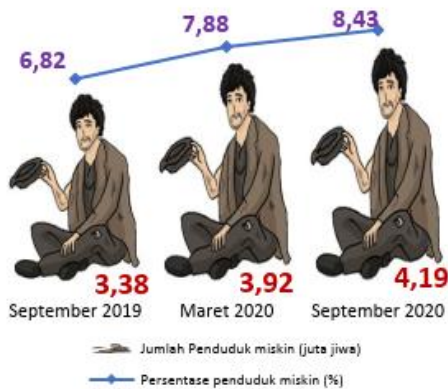
Berdasarkan latar belakang diatas maka untuk menentukan faktor-faktor yang berpengaruh terhadap tingkat kemiskinan dapat dilakukan dengan menggunakan metode analisis regresi. Namun, pada kenyataannya analisis regresi sederhana tidak tepat digunakan dikarenakan mengandung aspek spasial karena unit observasi dalam penelitian merupakan wilayah kabupaten/kota di Jawa Barat. Apabila hal ini diabaikan dapat menimbulkan taksiran yang bias dan tidak konsisten karena asumsi kebebasan antar unit observasi terlanggar (Anselin, 1988).

Karena keterbatasan data, maka dalam hal ini tidak semua variabel yang mempengaruhi tingkat kemiskinan dimasukkan ke dalam pemodelan seperti kebijakan pemerintah, sosial, budaya, pembangunan, dan lainnya yang memiliki aspek spasial sehingga akan dihitung sebagai error yang berkorelasi antar wilayah. Selain itu menurut Sugiyono (2003), pemerintah selalu melakukan pembangunan dengan memprioritaskan pembangunan ekonomi pada daerah perkotaan atau pusat perekonomian yang mengakibatkan daerah lainnya tertinggal dan kemiskinan menjadi tinggi. Hal tersebut menunjukkan faktor yang berhubungan dengan ekonomi diduga terdapat aspek spasial oleh karena itu maka menggunakan spasial lag variabel independen sebagai independen variabel.

Kemiskinan merupakan masalah multidimensial yang akan berakibat pada banyak sektor tentunya banyak faktor yang mempengaruhi tingkat kemiskinan yang tidak dimasukkan ke dalam pemodelan. Maka variabel independen yang tidak dimasukkan ke dalam pemodelan dinamakan omitted variables. Apabila omitted variables yang tidak dimasukkan ke dalam pemodelan dapat mengakibatkan taksiran yang bias dan tidak konsisten (Wooldridge, 2002). Dengan demikian, dalam mengetahui faktor-faktor kemiskinan di Jawa Barat diperlukan suatu pendekatan yang mampu mengakomodasi lag spasial independen variabel dan juga kekeliruan model yang berkorelasi spasial, serta mampu mengatasi bias taksiran akibat omitted variables

Maka dalam penelitian ini pendekatan metode Spatial Autoregressive Fixed Effect Model dalam mengidentifikasi faktor-faktor yang mempengaruhi tingkat kemiskinan di Jawa Barat. Tujuan dari penelitian ini adalah mendapatkan taksiran yang berpengaruh pada aktor tingkat kemiskinan pada dua puluh tujuh kabupaten/kota di Provinsi Jawa Barat sehingga menjadi dasar pertimbangan dalam pembuatan kebijakan bagi pemerintah daerah.

Berikut tabel tingkat kemiskinan Di Jawa Barat.



Gambar 1. Gambar Tingkat Kemiskinan Pada Tahun September 2019-2020.

Terlihat pada Gambar 1. bahwa Tingkat Kemiskinan menunjukkan hasil yang relatif masih tinggi maka diperlukan upaya dalam menanggulangi masalah kemiskinan yang terjadi.

2. METODE PENELITIAN

Data Hasil Penelitian

Pada penelitian yang dilakukan data yang digunakan didapatkan dari Badan Pusat Statistik Jawa Barat. Unit observasi meliputi 27 Kabupaten dan Kota di Provinsi Jawa Barat. Data penelitian yang digunakan adalah data dari tahun 2017 sampai tahun 2020 dengan rincian sebagai berikut :

- Y = Tingkat Kemiskinan
- X1 = Tingkat Pengangguran Terbuka (TPT)
- X2 = Indeks Pembangunan Manusia (IPM)
- X3 = Persentase Penduduk
- X4 = Tingkat Partisipasi Angkatan Kerja (TPAK)
- X5 = Produk Domestik Regional Bruto (PDRB).

Metode Analisis Data

Analisis Regresi

Analisis regresi yaitu metode dalam analisis statistik yang dapat digunakan dalam menjelaskan atau memodelkan hubungan antara variabel respon (Y) dengan satu atau banyak variabel prediktor (X1....Xp). Kegunaan analisis regresi diantaranya yaitu untuk memprediksi nilai variabel respon, dan melihat pengaruh satu atau lebih beberapa variabel prediktor terhadap variabel responnya. Terdapat beberapa pendekatan yang dapat digunakan dalam menaksir parameter pada model regresi, yaitu dengan menggunakan metode kuadrat terkecil (Ordinary Least Square), dan metode kemungkinan maksimum (MLE/Maximum Likelihood Estimate).

Regresi Spatial Data Panel

Menurut LeSage (1999), model regresi data panel yang diaplikasikan di beberapa wilayah terkadang Muhammad Irfan R¹, Teguh Ammar T²/
J Statistika Vol. 14, No. 1, (2021)

menghasilkan heterogenitas pengaruh error karena adanya dependensi dari unit cross section. Maka pengaruh spasial perlu dipertimbangkan dalam melakukan pemodelan yang melibatkan wilayah dan waktu.

Spatial Autoregressive/Lag Fixed Effect Model

Model SAR memiliki nilai $\rho \neq 0$, $\lambda = 0$, dan $\beta_w = 0$. Menurut Anselin (1988), persamaan model spasial lag data panel yaitu sebagai berikut:

$$y_{it} = \alpha_i + \rho \sum_{j=1}^N w_{ij}y_{jt} + \mathbf{x}_{it} \boldsymbol{\beta} + \mu_i + \varepsilon_{it} \quad (1)$$

Dalam notasi matriks dapat ditulis sebagai berikut:

$$\mathbf{y} = \alpha \mathbf{1}_N + \rho (\mathbf{I}_T \otimes \mathbf{W}_N) \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + (\mathbf{1}_T \otimes \mathbf{I}_N) \boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad (2)$$

$$\boldsymbol{\varepsilon} \sim \text{iid } N(0, \sigma^2 \mathbf{I}_n)$$

Spatial Error Fixed Effect Model

Model SEM memiliki nilai $\lambda \neq 0$, $\beta_w = 0$, dan $\rho = 0$. Menurut LeSage (1999), persamaan model spasial error data panel dapat dinyatakan sebagai berikut:

$$y_{it} = \alpha + \mathbf{x}_{it} \boldsymbol{\beta} + \mu_i + \varepsilon_{it} \quad (3)$$

$$\varepsilon_{it} = \lambda \sum_{j=1}^N w_{ij} \varepsilon_{jt} + \zeta_{it} \quad (4)$$

Dalam notasi matriks dapat ditulis sebagai berikut:

$$\mathbf{y} = \alpha + \mathbf{X} \boldsymbol{\beta} + (\mathbf{1}_T \otimes \mathbf{I}_N) \boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad (5)$$

$$\boldsymbol{\varepsilon} = \lambda \mathbf{W}_{NT} \boldsymbol{\varepsilon} + \boldsymbol{\zeta} \quad (6)$$

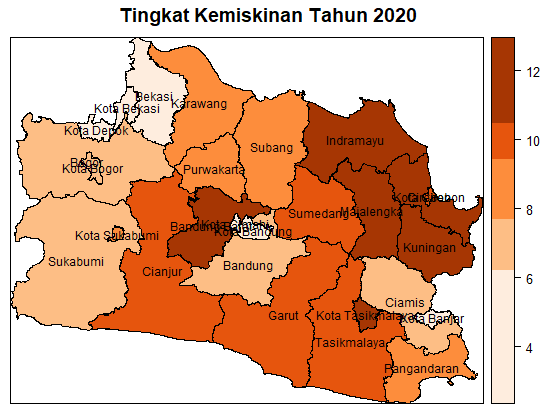
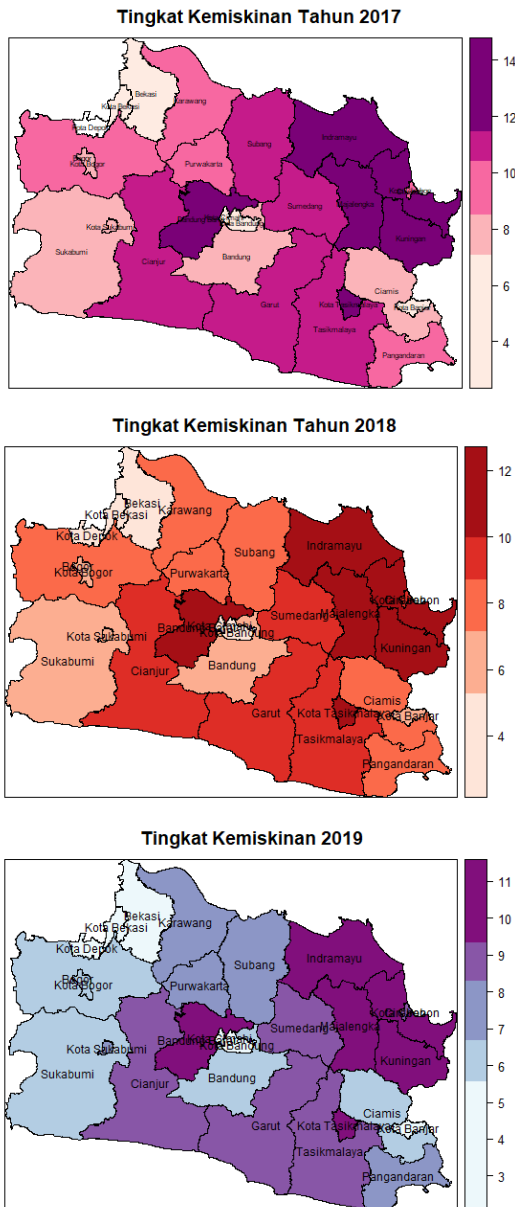
Langkah analisis regresi Spatial Autoregressive Fixed Effect Model.

1. Uji Aspek Spial untuk melihat ketergantungan antar wilayah dalam menggambarkan aspek spasial
2. Malakukan pem bobotan matriks Spaial dengan queen contiguity.
3. Uji Ketergantungan spasila dengan memperhatikan aspek wilayah terhadap waktu dengan menggunakan nilai dari Moran ST
4. Estimasi Parammeter dengan metode Kemungkinan maksimum (*Maximum Likelihood Estimator*/MLE)
5. Melakukan Pengujian Asumsi
 - a. Homoskedastisitas
 - b. Non-Autokorelasi

- c. Multikolinearitas
- d. Normalitas
- 6. Evaluasi model regresi spatial data panel
- 7. Goodness of Fit
- 8. Uji Signifikansi Parameter

3. HASIL DAN PEMBAHASAN

Berikut adalah hasil analisis regresi *Spatial Autoregressive Fixed Effect Model* dengan menggunakan taraf signifikansi sebesar 5%.



Gambar 2. Peta Penyebaran *Tingkat Kemiskinan* di Provinsi Jawa Barat.

Berdasarkan peta geografis pada gambar 2 diatas, semakin terang warna pada peta maka mengindikasikan bahwa semakin rendah suatu wilayah memiliki Tingkat Persentase kemiskinan yang tinggi. Kabupaten/Kota dengan warna yang semakin gelap menandakan bahwa kabupaten/Kota tersebut memiliki tingkat persentase tingkat kemiskinan tinggi. Berdasarkan peta penyebaran di Provinsi Jawa Barat selama empat tahun terakhir. Dapat terlihat bahwa Kota Tasikmalaya miliki secara terus-menerus selama empat tahun dan beberapa daerah dengan tingkat kemiskinan tinggi juga yaitu kuningan, Majalengka dan daerah yang berdekatan dengan wilayah tersebut.

Uji Aspek Spasial

Matriks Pembobot Spasial

Matriks pembobot spasial (**C**) *Queen Contiguity* dengan dimensi 27×27 adalah sebagai berikut:

$$= \begin{bmatrix} 0 & 0 & 1 & 0 & \dots & 1 \\ 0 & 0 & 1 & 0 & \dots & 1 \\ 1 & 1 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 0 & \dots & 0 \end{bmatrix}$$

Selanjutnya, nilai pada matriks **C** akan distandardisasi sehingga menghasilkan matriks pembobot spasial **W** sebagai berikut:

$$W = \begin{bmatrix} 0 & 1/8 & 1/8 & 0 & \dots & 0 \\ 1/3 & 0 & 1/3 & 0 & \dots & 0 \\ 1/6 & 1/6 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

Uji Ketergantungan Spasial

Dengan bantuan *software R.3.6.1 packages "MoranST"* diperoleh hasil perhitungan sebagai berikut:

Tabel 1. Pengujian Ketergantungan Spasial

<i>Moran's ST</i>	<i>p-value</i>
0.6294	0.0099

Berdasarkan output hasil Tabel 1 diatas, Dapat dilihat bahwa nilai indeks moran ST berada pada rentang $0 < I \leq 1$ artinya memperlihatkan adanya dependensi/autokorelasi spasial positif antar kecamatan. Pada Tabel 4.2 juga dapat dilihat *p-value* $\leq \alpha = 0.05$, maka H_0 ditolak. maka dengan taraf signifikansi 5% dapat disimpulkan terdapat dependensi/autokorelasi spasial antar Kabupaten atau kota di Jawa Barat. Maka dengan Adanya autokorelasi/ketergantungan spasial perlu diperhatikan dalam pembentukan dalam sebuah model.

1.1 Estimasi Parameter

Dengan bantuan *software R. 3.6.1 packages "spml"* maka diperoleh hasil sebagai berikut:

Tabel 2. Hasil Taksiran Parameter SAR-FEM

Parameter	<i>Estimate</i>	<i>Standard Error</i>	<i>p-value</i>
<i>Intercept</i>	8.6176	3.095	0.005363
Koefisien <i>Spatial Lag/Autoregressive</i> (ρ)	0.715018	0.049	0.0000
<i>Slope</i> variabel TPT	0.0745	0.0287	0.0095
<i>Slope</i> variabel IPM	-0.221	-4.772	0.000
<i>Slope</i> variabel Persentase Penduduk	1.928	0.549	0.00045
<i>Slope</i> variabel TPAK	0.0242	0.017	0.163
<i>Slope</i> variabel PDRB	0.001166	0.0038	0.7592
<i>Intercept</i> Kabupaten Bogor	-16.08	6.705	0.016
<i>Intercept</i> Kabupaten Sukabumi	-4.72	3.429	0.1685
<i>Intercept</i> Kabupaten Cianjur	-1.20	3.241	0.710

<i>Intercept</i> Kabupaten Bandung	-8.625	4.6215	0.0619
<i>Intercept</i> Kabupaten Garut	-3.153	3.53	0.371
<i>Intercept</i> Kabupaten Tasikmalaya	-0.31	2.923	0.914
<i>Intercept</i> Kabupaten Ciamis	0.215	2.786	0.93
<i>Intercept</i> Kabupaten Kuningan	5.530	2.651	0.036
<i>Intercept</i> Kabupaten Cirebon	-1.25	3.253	0.6995
<i>Intercept</i> Kabupaten Majalengka	3.159	2.6994	0.241
<i>Intercept</i> Kecamatan Sumur Bandung	3.255	2.773	0.240
<i>Intercept</i> Kabupaten Sumedang	1.99	2.941	0.496
<i>Intercept</i> Kabupaten Subang	0.565	2.90	0.84
<i>Intercept</i> Kabupaten Purwakarta	2.5244	2.675	0.345
<i>Intercept</i> Kabupaten Karawang	-1.425	3.452	0.679
<i>Intercept</i> Kabupaten Bekasi	-9.548	4.616	0.0386
<i>Intercept</i> Kabupaten Bandung Barat	2.726	2.915	0.349
<i>Intercept</i> Kabupaten Pangandaran	5.071	2.574	0.048
<i>Intercept</i> Kota Bogor	2.555	2.88	0.376
<i>Intercept</i> Koa Sukabumi	6.7637	2.696	0.0121
<i>Intercept</i> Kota Bandung	-4.264	3.805	0.262
<i>Intercept</i> Kota Cirebon	5.42	2.7337	0.04702
<i>Intercept</i> Kota Bekasi	-3.71	4.113	0.3670
<i>Intercept</i> Kota Depok	-4.16	3.67	0.25
<i>Intercept</i> Kota Cimahi	3.994	2.85	0.1616
<i>Intercept</i> Kota Tasikmalaya	9.486	2.69	0.0004
<i>Intercept</i> Kota Banjar	5.1929	2.652	0.050

Hasil taksiran parameter *Spatial Autoregressive/Lag Fixed Effect Model* yang tercantum pada tabel 2 diatas, menunjukkan bahwa Tingkat Pengangguran Terbuka (TPT) , Indeks Pembangunan Manusia (IPM) dan Persentase Penduduk memiliki pengaruh terhadap Tingkat Kemiskinan Di Provinsi Jawa Barat. Sedangkan Tingkat partisipasi Angkatan kerja (TPAK) dan Produk Domestik Regional Bruto tidak memiliki pengaruh terhadap Tingkat Kemiskinan Di Jawa Barat. Berdasarkan arah dari koefisien regresinya, dapat dilihat bahwa persentase IPM memiliki hubungan yang negatif terhadap Tingkat Kemiskinan. Sedangkan persentase Penduduk dan TPT memiliki hubungan yang positif terhadap Tingkat tingkat kemiskinan . Berdasarkan hasil taksiran parameter pada tabel 2 diatas, maka *Spatial Autoregressive/Lag Fixed Effect Model* (SAR-FEM) yang dapat terbentuk adalah sebagai berikut:

$$\hat{IR}_{it} = (\hat{\alpha} + \hat{\mu}_i) + 0.7150 \sum_{j=1}^N w_{ij} Y_{jt} + 0.074(TPT)_{it} - 0.221(IPM)_{it} + 1.928(PRS)_{it} \quad (36)$$

Pengujian Asumsi Homoskedastisitas

Kriteria uji untuk *Breusch-Pagan test* adalah tolak H_0 jika nilai $BP \geq \chi^2_{(k-1)}$, terima dalam hal lainnya. Berdasarkan *output* R, maka diperoleh nilai $BP = 4.50507 < \chi^2_{(4)} = 9.487$ maka H_0 diterima. Artinya, dengan tingkat kepercayaan 95% dapat disimpulkan bahwa varians *error* bersifat homoskedastisitas sehingga asumsi homoskedastisitas terpenuhi.

Non-Autokorelasi Spatio-Temporal

Pada pengujian non-autokorelasi *spatio-temporal* dapat dilakukan dengan menggunakan nilai Indeks Moran ST. Jika nilai MoranST mendekati satu, maka menunjukkan adanya autokorelasi *spatio-temporal* positif yang kuat dari kekeliruan model. Dengan bantuan *software R. 3.6.1 packages "MoranST"* maka diperoleh hasil yaitu nilai statistik uji MoranST sebesar -0.1708 dengan nilai $p - value = 0.970 > \alpha = 0.05$, maka H_0 diterima. Artinya, dengan tingkat kepercayaan 95% dapat disimpulkan bahwa tidak terdapat autokorelasi *spatio-temporal* pada kekeliruan model yang terbentuk untuk kasus Tingkat Kemiskinan di Provisni Jawa Barat.

Pendeteksian Multikolinearitas

Nilai determinan matriks yang diperoleh yaitu $det(X'X) = 1.7592 \times 10^{19}$. Dengan itu, dapat

disimpulkan tidak terdapat multikolinearitas sempurna. Selanjutnya, diperoleh nilai *Variance Inflation Factor* (VIF) yang digunakan dalam melihat ada atau tidaknya multikolinearitas antar variabel prediktor. Hasilnya yaitu sebagai berikut:

Tabel 3. Hasil Pedndeteksian Multikolinearitas

Variabel Prediktor	Nilai VIF
Tingkat Pengangguran Terbuka	1.515
Indeks Pembangunan Manusia	1.199
Persentase Penduduk	1.079
Tingkat Partisipasi Angkatan Kerja	1.365
Produk Domestik Regional Bruto	1.277

Berdasarkan Tabel 3, dapat dilihat bahwa ketiga variabel prediktor memiliki nilai *VIF* < 10 maka dapat disimpulkan bahwa tidak terdapat multikolinearitas antar variabel prediktor.

Normalitas

Pada asumsi normalitas digunakan uji *Jarque-Bera*. Kriteria uji untuk *Jarque-Bera test* adalah tolak H_0 jika nilai $JB \geq \chi^2_{(2)}$ atau $p - value \leq \alpha$, terima dalam hal lainnya. Dengan bantuan *software R. 3.6.1 packages "tseries"*, maka diperoleh nilai s nilai $p - value = 0.0039 < \alpha = 0.05$ maka H_0 ditolak. Artinya, dengan tingkat kepercayaan 95% dapat disimpulkan bahwa *error* tidak berdistribusi normal.

Karena sampel yang digunakan cukup besar yaitu sejumlah $(NxT) = 108$, maka tidak ada permasalahan yang timbul jika asumsi normalitas tidak terpenuhi. Penaksir OLS cenderung didistribusikan secara normal apabila banyaknya sampel meningkat secara tak terbatas. Ringkasnya, penaksir OLS dari koefisien regresi cenderung didistribusikan secara normal asimtotik. Namun hal tersebut tidak berlaku untuk jumlah sampel yang kecil (Gujarati, 2004).

Evaluasi Model

Lagrange Multiplier Test

Dengan bantuan *software R. 3.6.1 packages "bptest"* maka hasil yang diperoleh dari *LM test* yaitu sebagai berikut:

Tabel 4. Hasil LM Test

LM Test	Nilai Statistik Uji	p-value
LM-Joint	120.36	0.0000000+00
LM Marginal (1)	10.77	0.000

LM <i>Marginal</i> (2)	2.0513	4.02e-02
LM-rho <i>Conditional</i>	6.0846	1.16e-09
LM-mu <i>Conditional</i>	11.343	7.9e-30

Berdasarkan hasil dari tabel 4, kelima hipotesis tersebut yang memiliki *p-value* paling rendah yaitu hipotesis gabungan (a), maka dapat disimpulkan bahwa terdapat korelasi spasial atau terdapat keberagaman varians dari model yang dihasilkan.

Hausman Test

Hausman test digunakan untuk membandingkan model spasial lag *fixed effect* dengan model spasial lag *random effect*, dengan kriteria uji yaitu tolak H_0 jika nilai *p-value* $\leq \alpha$, terima dalam hal lainnya. Dengan bantuan *software R. 3.6.1 packages "sphtest"*, maka diperoleh hasil yaitu nilai *p-value* = 0.0005 < α = 0.05, maka H_0 ditolak. Artinya, dengan tingkat kepercayaan 95% dapat disimpulkan bahwa model yang sesuai yaitu *fixed effect model*.

Goodness of Fit

Pemodelan Tingkat Kemiskinan di Jawa Barat dengan menggunakan metode *spatial lag/autoregressive fixed effect model* menghasilkan nilai koefisien determinasi (R^2) sebesar 98.88%. Nilai tersebut memiliki arti bahwa model regresi yang terbentuk dapat menjelaskan besarnya keragaman dari Tingkat Kemiskinan yang dapat dijelaskan oleh variabel prediktor sebesar 98.88% sedangkan 1.116% sisanya dijelaskan oleh variabel lain yang tidak terlibat kedalam pemodelan.

Pengujian Signifikansi Parameter

Hasil signifikansi parameter dapat dilihat pada tabel sebagai berikut.

Tabel 5. Signifikasin Parameter

Parameter	Estimate	Standard Error	Nilai Wald	<i>p-value</i>
Koefisien <i>Spatial Lag</i> ($\hat{\rho}$)	0.762	0.049	14.5	0
<i>Slope</i> variabel TPT ($\hat{\beta}_{TPT}$)	0.075	0.028	2.60	0.004
<i>Slope</i> variabel IPM ($\hat{\beta}_{IPM}$)	-0.221	0.046	4.77	0

Muhammad Irfan R¹, Teguh Ammar T²/
J Statistika Vol. 14, No. 1, (2021)

<i>Slope</i> variabel PRS ($\hat{\beta}_{PRS}$)	1.928	0.550	3.51	0.00002
<i>Slope</i> variabel TPAK ($\hat{\beta}_{TPAK}$)	0.024	0.017	1.40	0.081
<i>Slope</i> variabel PDRB ($\hat{\beta}_{PDRB}$)	0.0011	0.003	0.306	0.379

Berdasarkan tabel 5 diatas, dapat dilihat bahwa parameter/koefisien spasial lag ρ memiliki nilai *p-value* < α = 0.05 maka H_0 ditolak. Dengan tingkat kepercayaan 95% dapat disimpulkan bahwa parameter spasial lag signifikan artinya terdapat dependensi spasial lag antar kabupaten dan kota yang berdekatan. Parameter β_{TPT} memiliki nilai *p-value* = 0.004 < α = 0.05 artinya H_0 ditolak, sehingga variabel tingkat pengangguran terbuka memberikan pengaruh yang signifikan dan parameter β_{IPM} dan Parameter β_{PRS} memiliki nilai *p-value* = 0.000 < α = 0.05 maka H_0 ditolak, artinya variabel IPM dan Persentase penduduk memberikan pengaruh yang signifikan terhadap tingkat kemiskinan. Sedangkan untuk parameter TPAK dan PDRB dengan alpha 5% tidak memberikan hasil signifikan terhadap tingkat kemiskinan di Jawa Barat.

4. KESIMPULAN

Berdasarkan hasil maka dapat ditarik beberapa kesimpulan, yaitu :

1. Variabel dalam penelitian yaitu tingkat pengangguran terbuka, indeks pembangunan manusia dan persentase penduduk berpengaruh signifikan terhadap Tingginya tingkat kemiskina di Provinsi Jawa Barat. hal ini sesuai dengan penelitian (Napitupulu, 2007) bahwa indeks pembangunan manusia mempunyai pengaruh dalam menurunkan tingkat kemiskinan. Selain itu tingginya sebuah kualitas sumber daya manusia tentunya akan berdampak kepada meningkatnya produktivitas tenaga kerja yang bisa meningkatkan perolehan pendapatan yang lebih tinggi. Dari pendapatan yang tinggi tentunya akan membuat masyarakat menjadi bisa dalam mencukupi kebutuhan hidupnya sendiri dan diharapkan bisa menurunkan tingkat kemiskinan (Sayifullah & Gandasari, 2016). Selain itu penelitian (Rochaida, 2016) menjelaskan bahwa ketika

jumlah penduduk dalam pembangunan ekonomi suatu wilayah merupakan permasalahan, diakibatkan karena pertumbuhan penduduk yang tidak terkendali akan menyebabkan tidak tercapainya tujuan pembangunan ekonomi yaitu kesejahteraan rakyat yang adil dan merata serta bisa menekan angka kemiskinan.

2. Model *spatial lag fixed effect* yang terbentuk dapat menjelaskan besarnya keragaman dari *Tingkat Kemiskinan* yang dapat dijelaskan oleh variabel prediktor sebesar 98.88% sedangkan 1.116% sisanya dijelaskan oleh variabel lain yang tidak dilibatkan kedalam pemodelan. Kabupaten dan Kota yang memiliki tingkat Kemiskinan paling rendah dalam kurun waktu tahun 2017-2020 yaitu Kota Depok dan dapat dilihat bahwa tingkat kemiskinan di beberapa daerah seperti kuningan, Majalengka dan daerah yang berdekatan dengan wilayah tersebut memiliki tingkat kemiskinan relatif tinggi. Maka dalam hal ini menjadi kebijakan khusus bagi pemerintah dalam mengetaskan permasalahan kemiskinan menjadi bahan pertimbangan dalam menentukan kebijakan pengendalian dan menekan tingkat kemiskinan dan juga dapat meningkatkan kesejahteraan masyarakat yang adil dan merata.

DAFTAR PUSTAKA

- Analisis Yan, X. & Xiao G.S. 2009. Linear Regression: Theory and Computing. World Scientific Publishing Co. Pte. Ltd.
- Anselin, L. 1988. Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers.
- Badan Perencanaan Pembangunan Daerah. 2018. Rancangan Pembangunan Jangka Menengah Daerah 2018-2023. Bandung: Bappeda Jawa Barat.
- Badan Perencanaan Pembangunan Nasional. 2004. Strategi Nasional Penanggulangan Kemiskinan. Jakarta: Bappenas
- Badan Pusat Statistik. (2016). Berita Resmi Statistik : Keadaan Ketenagakerjaan Indonesia Agustus 2020. Jakarta: Badan Pusat Statistik.
- Badan Pusat Statistik. (2017). Berita Resmi Statistik : Keadaan Ketenagakerjaan Indonesia Agustus 2020. Jakarta: Badan Pusat Statistik.
- Badan Pusat Statistik. (2018). Berita Resmi Statistik : Keadaan Ketenagakerjaan Indonesia Agustus 2020. Jakarta: Badan Pusat Statistik.
- Badan Pusat Statistik. (2019). Berita Resmi Statistik : Keadaan Ketenagakerjaan Indonesia Agustus 2020. Jakarta: Badan Pusat Statistik.
- Badan Pusat Statistik. (2020). Berita Resmi Statistik : Keadaan Ketenagakerjaan Indonesia Agustus 2020. Jakarta: Badan Pusat Statistik.
- Baltagi, B. H. (2005). Econometric Analysis of Panel Data (3th ed.). New York: John Wiley & Sons, Ltd.
- Draper, N.R. dan Smith, H. (1992). Applied Regression Analysis, Second Edition. New York: John Wiley and Sons.
- Harlik, A. A. (2013). Faktor-Faktor yang Mempengaruhi Kemiskinan dan Pengangguran di Kota Jambi . Jurnal Perspektif Pembiayaan dan Pembangunan Daerah Vol. 1 No. 2, Oktober 2013 , 1-12.
- Firman. (2019). Fixed Effect Panel Spatial Durbin Error Model Dalam Pengidentifikasian Faktor-Faktor Tingkat Kemiskinan Di Provinsi Jawa BARAT. Departemen Statistika Universitas Padjadjaran, 1-73.
- Gujarati, D. N. 2004. Basic Econometrics, Fourth Edition. New York: The Mc Graw-Hill Companies.
- Napitupulu, A. S. (2007). Pengaruh Indikator Komposit IPM Terhadap Penurunan Jumlah Penduduk Miskin Di Sumatera Utara.
- Rochaida, E. (2016). DAMPAK PERTUMBUHAN PENDUDUK TERHADAP PERTUMBUHAN EKONOMI DAN KELUARGA SEJAHTERA DI PROVINSI KALIMANTAN TIMUR. Forum Ekonomi.
- Sayifullah, & Gandasari, T. R. (2016). PENGARUH INDEKS PEMBANGUNAN MANUSIA DAN PENGANGGURAN TERHADAP KEMISKINAN DI PROVINSI BANTEN. JEQu, 236-255.
- Sugiyono, L. 2003. Karakteristik Kemiskinan dan Pemetaan Penduduk Miskin Provinsi Jawa Barat. Thesis: Institut Pertanian Bogor.
- Wooldridge, Jeffrey M. 2002. Econometric Analysis of Cross Section and Panel Data. London: The MIT Press.

Analisis Sentimen *Netizen Twitter* terhadap Pemberitaan PPN Sembako dan Jasa Pendidikan dengan Pendekatan *Social Network Analysis* dan *Naive Bayes Classifier*

Joko Ade Nursiyono ⁽¹⁾, Chusnul Chotimah ⁽²⁾

Badan Pusat Statistik Provinsi Jawa Timur

Jl. Raya Kendangsari Industri No 43-44 Telp. (031) 8439343 Surabaya 60292

e-mail: joko.ade@bps.go.id dan chus_chot@bps.go.id

ABSTRAK

Pandemi covid-19 yang terjadi memberikan dampak di berbagai bidang kehidupan. Salah satu dampaknya penerimaan negara semakin tertekan hebat. Padahal di sisi lain negara dalam proses pemulihan ekonomi nasional (PEN) yang membutuhkan dana sangat besar. Sehingga pemerintah ingin menggenjot pendapatan negara dari pajak pertambahan nilai (PPN). Jika pemungutan PPN dapat dilakukan dengan seoptimal mungkin, maka akan meningkatkan penerimaan negara. Rencana tersebut mengakibatkan maraknya pemberitaan mengenai pengenaan PPN sembako dan jasa pendidikan di Indonesia. Pemberitaan tersebut secara otomatis memicu opini di masyarakat. Salah satu cara untuk melihat opini masyarakat adalah melalui media sosial *Twitter*. Penelitian ini bertujuan untuk mengkaji *network* dan sentimen *netizen Twitter* tentang PPN Sembako dan jasa pendidikan menggunakan metode *betweenness centrality* dan *Naive Bayes Classifier*. *Social Network Analisis* (SNA) menghasilkan 5 kluster dengan *record* ke-90 merupakan *bottleneck node*, yaitu aktor utama penyebaran informasi antar kluster. Model *Naive Bayes Classifier* memberikan hasil *Accuracy Classified* sebesar 74.865 persen, sementara persentase *Incorrectly Classified Instance* sebesar 25.135 persen. Hasil klasifikasi berdasarkan emosi terbentuk 5 ekspresi *fear*, *sadness*, *surprise*, *joy*, dan *anger* dengan emosi kata terbanyak adalah emosi *anger* (amarah), artinya mayoritas respon masyarakat terhadap kebijakan pengenaan PPN sembako dan jasa pendidikan diidentifikasi oleh *R Studio* sebagai wujud keamarahan.

Kata kunci : *PPN, Sembako, Sentimen, Emosi Kata, Social Network Analisis, Naive Bayes Classifier.*

ABSTRACT

The COVID-19 pandemic has had an impact on many areas of life. One of the effects of state revenues is getting severely depressed. Whereas on the other hand, the country is in the process of national economic recovery (PEN) which requires very large funds. So the government wants to boost state revenue from value-added tax (VAT). If VAT collection can be carried out optimally, it will increase state revenue. The plan resulted in widespread news regarding the imposition of VAT on necessities and education services in Indonesia. The news automatically triggers public opinion. One way to see public opinion is through social media Twitter. This study aims to examine more deeply the network and sentiments of Twitter netizens about VAT for necessities and education services using betweenness centrality and Naive Bayes Classifier methods. The results of Social Network Analysis (SNA) resulted in 5 clusters with the 90th record being a bottleneck node, which is the main actor in disseminating information between clusters. The Naive Bayes Classifier model gives for Accuracy Classified it is 74,865 percent, while the percentage for Incorrectly Classified Instance is 25,135 percent. The results of the classification based on emotions formed 5 expressions of fear, sadness, surprise, joy, and anger and the emotion with the most words was anger, meaning that the majority of public responses to the policy of imposing VAT on necessities and education services were identified by R Studio as an anger emotion.

Keywords : *VAT, Basic Food, Sentiment, Word Emotion, Social Network Analysis, Naive Bayes Classifier.*

1. PENDAHULUAN

Pandemi covid-19 yang terjadi memberikan dampak di berbagai bidang kehidupan. Menurut Neilmardin (2021), salah satu dampaknya membuat penerimaan kas negara semakin tertekan hebat. Padahal di sisi lain negara dalam proses pemulihan

ekonomi nasional (PEN) yang membutuhkan dana sangat besar. Sehingga pemerintah ingin menggenjot pendapatan negara dari pajak pertambahan nilai (PPN). Tarif PPN Indonesia yang dipatok sebesar 10 persen terlalu rendah jika dibandingkan negara lain di dunia yang menerapkan

Pajak Pertambahan Nilai atau *Value Added Tax (VAT)* dan *Goods and Services Tax (GST)*. Struktur penerimaan negara dari PPN selama ini memiliki kontribusi sebesar 42 persen terhadap penerimaan negara. Jika pemungutan PPN dapat dilakukan dengan seoptimal mungkin, maka akan meningkatkan penerimaan negara sebagai bekal berbagai program pembangunan. Rencana tersebut mengakibatkan maraknya pemberitaan mengenai pengenaan PPN sembako dan jasa pendidikan di Indonesia. Pemberitaan tersebut secara otomatis memicu opini di masyarakat.

Opini yang terbentuk tentu menimbulkan sentimen sekaligus pro dan kontra di media sosial. Tingkat dan kecepatan penyebaran informasi ditentukan oleh seberapa besar pengaruh sebuah akun seseorang dalam jejaring media sosial, terutama *Twitter*. Melalui pendekatan *Social Network Analysis (SNA)*, penelitian ini bertujuan untuk menentukan aktor utama sekaligus pengaruhnya dalam model penyebaran informasi (Setatama, Mahdi, S. dan Tricahyono, Dodie, 2017) atau opini antar klaster. Selain itu, untuk mengetahui sentimen yang terbentuk dari wacana pengenaan PPN sembako maupun jasa pendidikan, penelitian ini menggunakan pengklasifikasian emosi kata model *Naive Bayes Classifier*. Menurut Nurzahputra, Aldi dan Muslim, Much A.(2016), dalam *text mining*, pendekatan *Naive Bayes Classifier* bisa digunakan untuk mengklasifikasikan opini ke dalam opini positif dan negatif.

2. METODE PENELITIAN

Set Data

Set data yang digunakan dalam penelitian ini bersumber dari *Twitter* yang diperoleh dengan teknik *Crawling* menggunakan *R Studio* versi 4.0.3. Dari sebanyak 5.000 target *records* yang akan digunakan sebagai sampel, didapatkan sebanyak 4.090 *records* yang terdiri atas 90 variabel. Adapun periode *records* yang tercatat sebagai sampel penelitian ini mulai 10 Juni 2021 - 21 Juni 2021. Setelah itu dilakukan pembersihan data teks (*text cleaning*) dengan tahapan seperti pada gambar 1.

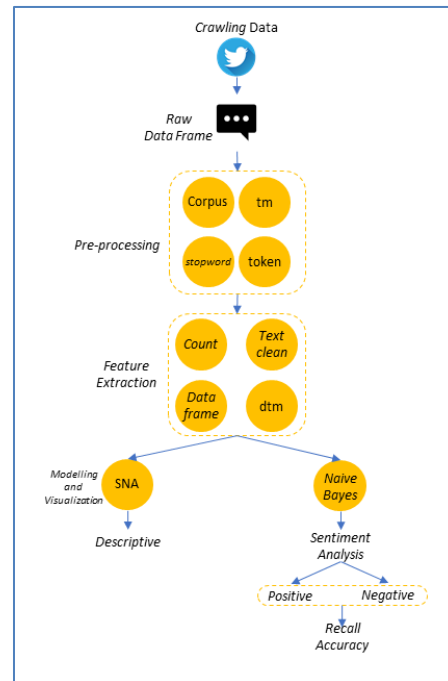
Import data

Import data dilakukan dengan menggunakan library *readxl* kemudian menyeleksi kalimat *tweet* pada data untuk kemudian dijadikan sebagai data *corpus* melalui library *corpus*.

Text Cleaning

Pada data *corpus*, dilakukan pembersihan kalimat dari tautan atau URL, *mention*, *hashtags* atau #,

menghapus kelebihan spasi, dan nomor menggunakan library *tm*.



Gambar 1. Bagan pembersihan data teks (*text cleaning*)

Stopwords Bahasa Indonesia

Pada struktur data yang sama, untuk menghapus kata hubung (konjungsi) atau kata-kata lain yang tidak bermakna, termasuk kata singkatan dilakukan *tokenize* kemudian mengeliminasi dengan fungsi *custome*, yaitu *stopwords-Indonesia*.

Export Clean Text

Setelah ketiga tahapan di atas dilakukan, pada tahap akhir adalah mengubah kembali *tweets* dalam bentuk *data frame* dan menyimpannya dalam *file excel* dengan nama: *databersih*.

Analisis Jaring Sosial (Social Network Analysis)

Social Network Analysis (SNA) merupakan salah satu analisis untuk melihat tingkat keeratan dan pengklasifikasian topik yang dibahas dalam jejaring sosial. Menurut Alamsyah, Andri dan Bratawisnu, Made K. (2018), analisis yang dapat dilakukan terkait jejaring sosial yang terbentuk adalah properti jaringan sosial. Properti sosial yang terbentuk berupa *nodes*, *edges*, dan *degree* untuk selanjutnya dianalisis *Centrality*-nya.

Tabel 1. Pengertian properti jaringan

Properti Jaringan	Definisi
<i>Nodes</i>	<i>Nodes</i> adalah aktor individu dalam jaringan yang dihubungkan oleh <i>ties</i> (Oktora, Rio dan Alamsyah, Andry, 2014)
<i>Edges</i>	<i>Edges</i> merepresentasikan interaksi dua aktor (<i>nodes</i>) atau lebih (Anam, Khairul M, dkk., 2021)
<i>Degree</i>	<i>Degree</i> merupakan ukuran keeratan grup di dalam jaring (<i>Network</i>)
<i>Centrality</i>	<i>Centrality</i> adalah proses kalkulasi terhadap relasi setiap individu dari sebuah jejaring sosial (Susanto, Budi. dkk., 2012)

Centrality

Centrality pada dasarnya menentukan *node* utama penyebaran sebuah informasi. Proses kalkulasi *centrality* dapat dihitung dengan pendekatan jumlah interaksi atau *edge* antar *node*, jarak terpendek yang ditempuh oleh sebuah *node* ke seluruh *node*, berdasarkan skor *eigen vector* jaring antar *node*, dan ukuran peran sebuah *node* terhadap seluruh *node*. Untuk menentukan *centrality*, dalam penelitian ini menggunakan pendekatan *betweenness centrality* yang merupakan modifikasi dari *degree centrality*. Menurut Tabassum et al. (2018), *degree centrality* adalah jumlah interaksi atau *edge* yang terjadi pada suatu *node* tertentu dengan menggunakan formula berikut:

$$C_D(i) = \sum_{j=1}^n a_{ij} \tag{1}$$

di mana:

C_D : *Degree centrality node ke-i*

a_{ij} : nilai baris ke-*i* dan kolom ke-*j* matriks *adjacency*

Degree centrality ditentukan oleh matriks *adjacency*. Menurut Santoso, Budi. dkk. (2012), nilai yang ada di setiap *cell* matriks *adjacency* menunjukkan informasi atas hubungan atau relasi antar aktor atau individu. Menurut Mamas, Christoporos (2018), struktur matriks *adjacency* untuk menentukan *degree centrality Social Network Analysis* (SNA) adalah sebagai berikut:

	Node 1	Node 2	Node 3	Node 4	Node 5
Node 1	—	1	1	0	0
Node 2	1	—	0	0	0
Node 3	0	0	—	1	1
Node 4	0	1	0	—	1
Node 5	0	1	1	1	—

Gambar 2. Struktur matriks *adjacency*

Betweenness Centrality

Betweenness centrality pada dasarnya merupakan ukuran untuk melihat peranan sebuah *node* dalam jaring atau *network*. Menurut Pratama, Faustinus Y. (2018), *betweenness centrality* dapat digunakan untuk melihat *node* yang berperan sebagai *bottleneck*. Sebuah *node* dikatakan mempunyai peran penting jika menjadi *communication bottleneck*. Ukuran ini juga dapat digunakan untuk mengidentifikasi *boundary spanners*, yaitu aktor atau *node* yang berperan sebagai penghubung (*bridge*) antara komunitas. Ukuran ini dihitung dengan menjumlahkan semua garis atau *path* yang mengandung *bottleneck node*. Rumus di bawah ini digunakan dalam menghitung *betweenness centrality*:

$$C_i^B = \frac{1}{(N-1)(N-2)} \sum_{j \in G, G \neq 1} \sum_{k \in G, k \neq i, k \neq j} \frac{n_{jk}(i)}{n_{jk}} \tag{2}$$

di mana:

C_i^B : bobot nilai *betweenness centrality*

N : jumlah *node* dalam suatu jaringan

n_{jk} : jumlah jalur (*path*) terpendek dari *node ke-j* ke *node ke-k*

$n_{jk}(i)$: jumlah jalur (*path*) terpendek dari *node ke-j* ke *node ke-k* yang melewati simpul ke-*i*

Analisis Sentimen

Analisis sentimen merupakan *tool* analisis untuk data yang bertipe teks dan kontekstual. Menurut Rozi, Imam F, dkk. (2012), analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini.

Naive Bayes Classifier

Naive Bayes Classifier (NBC) merupakan algoritma pengklasifikasian yang sering digunakan dalam permasalahan teks (Rozi, Imam F, dkk., 2012). Adapun atribut yang digunakan dalam NBC adalah data *training*, *prior probability*, serta *posterior probability*. Untuk menjelaskan teori *Naive Bayes*, proses pengklasifikasian memerlukan sejumlah

petunjuk untuk menentukan kelas yang cocok bagi sampel yang dianalisis. Secara umum, peluang teorema *Bayes* yang disesuaikan dirumuskan berikut:

$$P(C|F_1, F_2, F_3, \dots, F_n) = \frac{P(C) \cdot P(F_1, F_2, F_3, \dots, F_n|C)}{P(F_1, F_2, F_3, \dots, F_n)} \quad (3)$$

di mana C merepresentasikan kelas dan $F_1, F_2, F_3, \dots, F_n$ merupakan karakteristik yang diperlukan sebagai dasar pengklasifikasian. $P(C|F_1, F_2, F_3, \dots, F_n)$ pada rumus tersebut merupakan *posterior probability*, yaitu peluang masuknya sampel karakteristik tertentu dalam kelas C . Sedangkan $P(C)$ merupakan *prior probability*, yaitu besarnya peluang munculnya kelas C sebelum sampel tersebut masuk. $P(F_1, F_2, F_3, \dots, F_n|C)$ merupakan peluang munculnya karakter-karakter sampel kelas C . Dan $P(F_1, F_2, F_3, \dots, F_n)$ disebut sebagai peluang munculnya karakter sampel secara global atau disebut *evidence*. Setelah sentimen untuk masing-masing *record* terbentuk, kebaikan model *Naive Bayes Classifier* dalam penelitian ini diukur menggunakan *Recall Accuracy*. Menurut Rozi, Imam F, dkk. (2012), *Recall* merupakan rasio jumlah dokumen yang relevan yang ditemukan terhadap total jumlah dokumen relevan yang dikumpulkan. Nilai *recall* biasanya ditunjukkan dalam satuan persen (%). Ukuran ini merupakan *proxy* performa model *Naive Bayes Classifier* dalam mengevaluasi kecocokan pengklasifikasian sentimen seluruh sampel. Berikut ini adalah rumus penghitungan *recall accuracy* (Handayani, Fitri dan Pribadi, Feddy S., 2015) :

$$Recall Accuracy = \frac{TP}{TP + FP} \quad (4)$$

di mana:

TP : klasifikasi sentimen yang benar sesuai model *Naive Bayes*

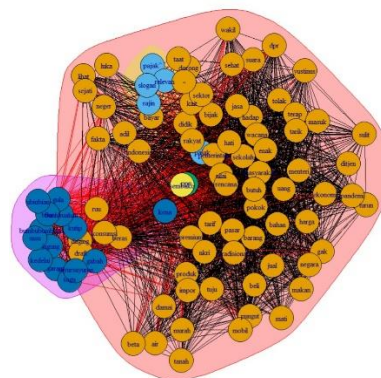
FP : klasifikasi sentimen yang salah sesuai model *Naive Bayes*

Berdasarkan nilai *recall accuracy*, semakin besar nilai persentasenya, maka menunjukkan hasil pengklasifikasian model *Naive Bayes Classifier* semakin baik.

3. HASIL DAN PEMBAHASAN

Hasil pengolahan dengan menggunakan *Social Network Analysis* (SNA) terbentuk 5 klaster. Klaster 1 memiliki 73 anggota, klaster 2 memiliki 5 anggota, klaster 3 memiliki 1 anggota, dan klaster 4 memiliki 1 anggota, serta klaster 5 memiliki 14 anggota.

Hasil dari *centrality summary* didapatkan nilai minimum sebesar -0.028, kemudian kuantil pertama sebesar 0.010, dan nilai maksimumnya sebesar 0.073. Berdasarkan *summary* tersebut, nilai maksimum *centrality* terdapat pada *record* ke-90 dan sekaligus merupakan *bottleneck node*, yaitu aktor utama penyebaran informasi antar klaster dengan kalimat *tweet*, “Pemerintah berencana mengenakan Pajak Pertambahan Nilai (PPN) terhadap sejumlah kebutuhan masyarakat banyak, termasuk di antaranya sembako dan jasa pendidikan (sekolah). Apa muara dari permasalahan ini? Simak analisis selengkapnya.” Dari 5 klaster yang terbentuk, *record* ke-90 ini masuk dalam klaster sendiri, yaitu klaster 4 (terlihat *node* warna kuning pada gambar 3).



Gambar 3. Hasil olah *social network analisis* (SNA)

Analisis Sentimen

Twitter merupakan salah satu media sosial yang biasa digunakan untuk melihat opini masyarakat. Berdasarkan hasil olah data dari *twitter*, terdapat beberapa kata yang paling banyak muncul dalam *tweet* terkait isu PPN sembako dan jasa pendidikan. Kata yang sering digunakan yaitu “pajak”, “ppn”, “pemerintah”, “kena”, “rakyat”, dan tentunya masih banyak kata-kata juga digunakan dalam *tweet*. Artinya, secara umum masyarakat menggunakan satu kata dari kelima kata tersebut dalam unggahan *tweet*-nya terkait kebijakan pengenaan PPN sembako dan jasa pendidikan (lihat gambar 4).

Classifier dalam penelitian ini diukur menggunakan *Recall Accuracy*. Hasil penghitungan *Recall*

Accuracy menunjukkan bahwa untuk *Accuracy Classified* sebesar 74.865 persen sementara persentase untuk *Incorrectly Classified Instance* sebesar 25.135 persen. Pada gambar 7, terlihat bahwa dari 4090 data *twitter* terkait topik ppn sembako dan jasa pendidikan, ada sebanyak 3062 data yang berhasil diklasifikasikan sentimennya dengan benar dan sebanyak 1028 data *twitter* lainnya tidak berhasil diklasifikasikan dengan benar.

Pada tabel 2 dapat dilihat bahwa dari 413 *tweet* yang diklasifikasikan bersentimen negatif ternyata semuanya masuk sentimen netral, dari 615 *tweet* bersentimen positif, ternyata semuanya masuk klasifikasi sentimen netral, hanya sebanyak 3062 *tweet* saja yang diklasifikasikan menurut sentimen yang benar (yaitu netral).

Tabel 2. Klasifikasi prediksi sentimen model *naive bayes*

Prediksi	Negatif	Netral	Positif
Negatif	0	0	0
Netral	413	3062	615
Positif	0	0	0

Sumber: hasil olah *R Studio* versi 4.0.3

4. KESIMPULAN DAN SARAN

1. Pengolahan data tweet ppn sembako menggunakan Social Network Analisis (SNA) menghasilkan 5 kluster dengan record ke-90 merupakan bottleneck node yaitu aktor utama penyebaran informasi antar kluster.
2. Hasil analisis sentimen secara umum masyarakat menggunakan satu kata dari kelima kata yang sering digunakan dalam tweet terkait pemberitaan ppn sembako dan jasa pendidikan yaitu pemerintah, pajak, ppn, didik, rakyat dan tentunya masih banyak kata-kata juga digunakan dalam tweet.
3. Hasil penghitungan Recall Accuracy menunjukkan bahwa untuk Accuracy Classified sebesar 74.865 persen sementara persentase untuk Incorrectly Classified Instance sebesar 25.135 persen.
4. Hasil klasifikasi berdasarkan emosi terbentuk 5 ekspresi fear, sadness, surprise, joy, dan anger dan emosi kata yang paling banyak adalah emosi anger (amarah), artinya, mayoritas respon masyarakat terhadap kebijakan pengenaan PPN sembako dan jasa pendidikan diidentifikasi oleh R Studio sebagai wujud keamaraan.

5. Kendati masih wacana, ternyata di media sosial Twitter sudah terbentuk sentimen. Dengan demikian, sebaiknya pengumuman kebijakan PPN sembako dan jasa pendidikan diumumkan secara detil agar tidak menimbulkan salah persepsi di masyarakat.

DAFTAR PUSTAKA

- Anam, Khairul M., dkk. 2021. Analisis Kesiapan Masyarakat Pada Penerapan Smart City di Sosial Media Menggunakan SNA, *Jurnal Rekayasa Sistem dan Teknologi Informasi*, 5(1), 69-81.
- Alamsyah, Andry. dan Bratawisnu, Made. K. 2018. Social Network Analysis Untuk Analisa Interaksi User Media Sosial Mengenai Bisnis E-Commerce (Studi Kasus: Lazada, Tokopedia, dan Elevenia). *Jurnal Manajemen dan Bisnis*, 2(2), 107-115.
- DJP Buka 3 Alasan Pemerintah Pungut PPN Sembako dan Sekolah. 2021. <https://www.cnnindonesia.com/ekonomi/20210614101335-532-653968/djp-buka-3-alasan-pemerintah-pungut-ppn-sembako-dan-sekolah>, diakses tanggal 21 Juni 2021.
- Handayani, Fitri dan Pribadi, Feddy, S. 2015. Implementasi Algoritma Naïve Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110, *Jurnal Teknik Elektro*, 7(1), 19-24.
- Mamas, Christoforos, 2018. The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation, Chapter Title: "Matrices (in Social Network Analysis)", SAGE Publication, Inc, Thousand Oaks.
- Nurzahputra, Aldi dan Muslim, Much A. 2016. Analisis Sentimen pada Opini Mahasiswa Menggunakan Natural Language Processing, *Seminar Nasional Ilmu Komputer (SNIK 2016)-Semarang*, 10 Oktober 2016, 114-118.
- Oktora, Rio dan Alamsyah, Andy. 2014. Pola Interaksi dan Aktor yang Paling Berperan pada Event JGTC 2013 melalui Media Social Twitter (Studi menggunakan Metode Sosial Networking Analysis), *Jurnal Manajemen Indonesia*, 14(3), 201-209.

Pratama, Faustinus Y. 2018. Simulasi Jejaring Jalan Kota Pontianak Dengan Betweenness Centrality dan Degree Centrality, Jurnal TIN Universitas Tanjungpura, 2(1), 1-6.

Rozi, Imam F, dkk. 2012. Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi, Jurnal EECCIS, 6(1), 37-43.

Setatama, Mahdi, S. dan Tricahyono, Dodie, 2017. Implementasi Social Network Analysis dalam Penyebaran Country Branding “Wonderful Indonesia”, 2(1), 91-104.

Susanto, Budi, dkk. 2012. Penerapan Network Analysis dan Penentuan Centrality Studi Kasus Social Network Twitter, Jurnal Informatika, 8(1), 1-13.

Tabassum, Shazia, dkk. 2018. Social Network Analysis: an Overview, Original Article Jurnal Sectoin.

JStatistika : Jurnal Ilmiah Teori Dan Aplikasi Statistika ber ISSN online 2654-7511 dan ISSN cetak 2089-0028,

JStatistika merupakan media penerbitan karya ilmiah berupa literasi hasil penelitian ataupun kajian tentang teori, metodologi serta aplikasi dalam bidang ilmu statistika dan data sains. Jstatistika diterbitkan oleh Program Studi Statistika Fakultas Sains dan Teknologi Universitas PGRI Adi Buana Surabaya dan akan terbit dua kali dalam setahun setiap bulan juli dan desember.

**FAKULTAS
SAINS DAN TEKNOLOGI**
Universitas PGRI Adi Buana Surabaya

