ISSN 2623-0658 (print)
ISSN 2623-2286 (on-line)

Volume 2, Number 2, December 2019

Journal of Data Analysis

Statistics and Its Applications

Published by:

Department of Statistics
Faculty of Mathematics and Natural Sciences
Universitas Syiah Kuala
Banda Aceh

http://jurnal.unsyiah.ac.id/JDA

Editorial Team

Journal of Data Analysis: Statistics and its Application,

Editor Address

Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala
Jalan Syech Abdurrauf No.3, Kopelma Darussalam, Banda Aceh 23111, Aceh, Indonesia
Email: j.data.analysis@gmail.com, Website: http://jurnal.unsyiah.ac.id/jda
Diterbitkan dua kali dalam setahun

Journal of Data Analysis: Statistics and Its Applications
Preface

Assalamu'alaikum Wr Wb.

Alhamdulillah, Journal of Data Analysis (JDA): Statistics and Its Applications Volume 2 Number 2 December 2019 can be published. This edition consists of 6 (six) articles which are mathematics, informatics and applied statistics.

We hope that the journal can be useful to readers and statisticians for developing knowledge in statistics and mathematics subjects not only theory but also applications. The subjects of the articles are game theory, assurance, design of experiment, neural network, algebra group, cluster analysis and fuzzy system.

The journal is made by including many people therefore we thank to authors, reviewers, and the others which have helped from preparation until finishing the journal.

Wassalamu'alaikum Wr Wb.

Banda Aceh, 29 May 2020

Editorial Team

ISSN 2623-0658 (Print)
ISSN 2623-2286 (On-line)

Journal of Data Analysis
Statistics and Its Applications

Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala
Jalan Syech Abdurrauf No.3, Kopelma Darussalam, Banda Aceh 23111, Aceh, Indonesia
Email: j.data.analysis@gmail.com, Website: http://jurnal.unsyiah.ac.id/jda

Table of Contents

Journal of Data Analysis
Vol.2, No.2, December 2019, p. 53 - 60

Beberapa Subgrup dari $SL(2,3)$

Mahmudi[1*], Ikhsan Maulidi[2], Saiful Amri[3]

[1,2,3] Department of Mathematics, Universitas Syiah Kuala, Banda Aceh, Indonesia
E-mail: mahmudi@unsyiah.ac.id*

Abstrak

Artikel ini membahas mengenai $SL(2,3)$ dengan rincian elemen-elemennya. Dengan bantuan Tabel Cayley dibuktikan bahwa $SL(2,3)$ merupakan grup dan memiliki beberapa subgrup siklik dan subgrup tidak siklik sesuai dengan Teorema Lagrange. Lebih jauh, juga dibuktikan bahwa $SL(2,3)$ tidak memiliki subgrup berorder 12.

Informasi Artikel

Sejarah Artikel:
Diajukan 9 Feb, 2020
Diterima 21 Feb, 2020

Kata Kunci:
Grup $SL(2,3)$
Subgrup Siklik
Tabel Cayley
Teorema Lagrange

Abstract

This article discusses about $SL(2,3)$ with its detail elements. By using Cayley Table, we prove $SL(2,3)$ is a group and has cyclic subgroup and noncyclic subgroup according to The Lagrange Theorem. Futher, we also give a detail proof that $SL(2,3)$ has no subgroup of order 12.

Keyword:
Group $SL(2,3)$
Cyclic Subgroup
Cayley Table
Lagrange Theorem

1. Pendahuluan

Grup permutasi berderajat n, yaitu S_n, memiliki subgrup dengan elemen berupa permutasi genap yang dinamakan grup alternating berderajat n, dinotasikan A_n [1]. Grup alternating merupakan contoh sangkalan bahwa konvers Teorema Lagrange tidak berlaku. Gallian membuktikan pada grup A_4 yang berorder 12, tidak memiliki subgrup berorder 6, Padahal diketahui bahwa bilangan 6 merupakan salah satu pembagi dari 12 [1]. Brennan dan Machale memberikan beberapa versi bukti terkait dengan konvers Teorema Lagrange pada A_4 [2].

Contoh-contoh yang terkait konvers Teorema Lagrange juga dikenal dengan subgrup berindeks dua, Nganou menyatakan bahwa contoh-contoh tersebut sangat jarang [3]. Mackiw membahas persoalan konvers Teorema Lagrange pada grup $SL(2,3)$ [4], [5]. Grup $SL(2,3)$ merupakan grup multiplikatif dengan elemen berupa matriks berordo 2×2 dengan entri-entri merupakan elemen lapangan $\mathbb{Z}_3$.

Berdasarkan pembahasan Mackiw dan beberapa ide yang ditulis oleh Shariari melalui soal yang belum diselesaikan dalam bukunya [6], artikel ini membahas beberapa detail terkait grup $SL(2,3)$ dengan menggunakan bantuan Tabel Cayley. Artikel ini akan menyajikan detail elemen-elemen di $SL(2,3)$, kemudian menampilkan hasil operasi elemen tersebut dalam Tabel Cayley. Artikel ini juga membahas cara menentukan beberapa subgrup siklik dan subgrup tidak siklik dari $SL(2,3)$. Pada bagian akhir, dengan melengkapi berbagai detail bukti yang diklaim oleh Mackiw, penulis membuktikan bahwa $SL(2,3)$ tidak memiliki subgrup berorder 12.

2. Tinjauan Kepustakaan

Secara umum pembaca diasumsikan telah mengenal definisi grup dan beberapa sifat dasarnya. Detail dapat dipelajari pada Gallian [1], Gilbert dan Gilbert [7], dan Hungerford [8]. Grup dengan banyaknya elemen berhingga dinamakan grup hingga [8], banyaknya elemen grup berhingga dinamakan order grup G, dinotasikan $o(G)$ [1]. Grup komutatif adalah grup dengan sifat jika hasil operasi dengan menukarkan posisi elemen pertama dengan elemen kedua memberikan hasil yang sama, secara matematis ditulis $a * b = b * a$ untuk setiap $a, b \in G$ [1].

Subgrup didefinisikan sebagai himpunan tak kosong H di G yang membentuk grup terhadap operasi yang sama di G [7]. Untuk sebarang grup G, himpunan semua elemen yang bersifat komutatif terhadap semua elemen di G dinamakan pusat (*center)* dari G, yang dinotasikan $Z(G)$ [7]. Secara matematis dituliskan sebagai

$$Z(G) = \{a \in G | ax = xa \text{ untuk setiap } x \in G\}.$$

Dapat dibuktikan bahwa $Z(G)$ merupakan subgrup dari G [1].

Salah satu subgrup yang menjadi fokus pembahasan dalam artikel ini adalah subgrup siklik. Misalkan x adalah elemen di G, maka subgrup siklik adalah subgrup yang dibangun oleh x elemen G, secara matematis dinotasikan sebagai $\langle x \rangle = \{a \in G | a = x^n, \text{untuk suatu } n \in Z\}$ [7].

Dengan memiliki subgrup maka dapat dibentuk himpunan baru yang dinamakan dengan koset. Misalkan H subgrup dari G dan $x \in G$, maka $xH = \{xh | h \in H\}$ dinamakan koset kiri dari H di G. Dengan cara serupa, $Hx = \{hx | h \in H\}$ dinamakan koset kanan dari H di G [6]. Mahmudi membahas dengan detail cara pembentukan koset dan operasi pada himpunan koset agar terdefinisi dengan baik [9] sehingga dapat membentuk grup.

Salah satu teorema penting terkait pembahasan artikel ini adalah Teorema Lagrange yang dapat dituliskan sebagai berikut.

Teorema Lagrange [1]. Misalkan G adalah grup berhingga dan H subgrup dari G maka orde H membagi habis order G. Lebih jauh, banyaknya koset kiri (kanan) dari subgrup H di grup G adalah $\frac{o(G)}{o(H)}$. Bilangan ini dinamakan dengan indeks H di G [1].

Bukti. Bukti Teorema Lagrange dapat dibaca pada Gallian [1].

Konvers Teorema Lagrange dapat dinyatakan misalkan G grup berhingga berorde n, jika bilangan k pembagi bilangan n maka G memiliki subgrup berorder k. Telah banyak contoh sangkalan yang disajikan untuk membantah Konvers Teorema Lagrange. Dalam artikel ini, grup $SL(2,3)$ yang merupakan grup berhingga dengan order 24 dibuktikan tidak memiliki subgrup berorder 12, yang diketahui bahwa bilangan 12 merupakan pembagi bilangan 24.

3. Hasil dan Pembahasan

Diberikan lapangan $\mathbb{Z}_3 = \{0, 1, 2\}$ maka dapat dibentuk himpunan

$$SL(2,3) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| a, b, c, d \in \mathbb{Z}_3, ad - bc = 1 \right\}.$$

Pembahasan pertama terkait dengan order $SL(2,3)$, yaitu akan dihitung dengan rinci banyaknya elemen $SL(2,3)$. Perhitungan tersebut dapat dilakukan dengan cara sebagai berikut.

Kasus 1. Jika $ad = 0$ maka haruslah $bc = 2$, supaya determinan matriks

$$ad - bc = 0 - 2 = -2 \equiv 1.$$

Dengan demikian, $ad = 0$ hanya dipenuhi jika $a = 0$ atau $d = 0$. Jika $a = 0$ akan diperoleh bentuk matriks

$$\begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 2 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 2 & 2 \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ 1 & 2 \end{pmatrix}.$$

Jika $d = 0$, cukup menukar posisi bilangan pada diagonal utama dan membuang dua matriks yang berulang, akan diperoleh matriks yang sama sehingga diperoleh bentuk-bentuk

$$\begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 2 \\ 1 & 0 \end{pmatrix}.$$

Artinya, terdapat 10 elemen matriks dengan $ad = 0$ dan $bc = 2$.

Kasus 2. Berikutnya, jika hasil perkalian entri diagonal sekunder bernilai 0, yaitu $bc = 0$, maka haruslah $ad = 1 \equiv 4$, supaya determinan matriks

$$ad - bc = 1 - 0 = 1$$

atau

$$ad - bc = 4 - 0 = 4 \equiv 1.$$

Dengan demikian, $ad = 1 \equiv 4$ hanya terjadi, jika $a = d = 1$ atau $a = d = 2$, sehingga diperoleh

bentuk-bentuk

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix},\begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix},\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix},\begin{pmatrix}1 & 0\\ 2 & 1\end{pmatrix},\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix},\begin{pmatrix}2 & 1\\ 0 & 2\end{pmatrix},\begin{pmatrix}2 & 2\\ 0 & 2\end{pmatrix},\begin{pmatrix}2 & 0\\ 1 & 2\end{pmatrix},\begin{pmatrix}2 & 0\\ 2 & 2\end{pmatrix}.$$

Kedua kasus tersebut merupakan bentuk trivial karena salah satu diagonalnya bernilai nol. Selain itu, juga terdapat kemungkinan diagonal matriks tidak bernilai nol. Beberapa kemungkinan tersebut dibahas pada kasus 3 berikut.

Kasus 3. Nilai determinan matriks 1 dapat juga diperoleh dengan kombinasi $ad = 2$ dan $bc = 1$, yaitu

$$ad - bc = 2 - 1 = 1$$

atau $ad = 2$ dan $bc = 4$, yaitu

$$ad - bc = 2 - 4 = -2 \equiv 1.$$

Dengan demikian, akan diperoleh bentuk-bentuk matriks

$$\begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix},\begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix},\begin{pmatrix}1 & 2\\ 2 & 2\end{pmatrix},\begin{pmatrix}2 & 2\\ 2 & 1\end{pmatrix}.$$

Berdasarkan 3 kasus yang telah diuraikan, diperoleh total elemen $SL(2,3)$ sebanyak $10 + 10 + 4 = 24$ elemen. Detail elemen-elemenya dituliskan dalam indeks sebagai berikut.

$$\begin{aligned}SL(2,3) = \Big\{&a_1 = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\ a_2 = \begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}, a_3 = \begin{pmatrix}0 & 1\\ 2 & 2\end{pmatrix}, a_4 = \begin{pmatrix}0 & 2\\ 1 & 2\end{pmatrix}, a_5 = \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix},\\ &a_6 = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, a_7 = \begin{pmatrix}1 & 0\\ 2 & 1\end{pmatrix}, a_8 = \begin{pmatrix}2 & 1\\ 2 & 0\end{pmatrix}, a_9 = \begin{pmatrix}2 & 2\\ 1 & 0\end{pmatrix}, a_{10} = \begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix},\\ &a_{11} = \begin{pmatrix}0 & 1\\ 2 & 0\end{pmatrix}, a_{12} = \begin{pmatrix}0 & 2\\ 1 & 0\end{pmatrix}, a_{13} = \begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}, a_{14} = \begin{pmatrix}1 & 2\\ 2 & 2\end{pmatrix}, a_{15} = \begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix},\\ &a_{16} = \begin{pmatrix}2 & 2\\ 2 & 1\end{pmatrix}, a_{17} = \begin{pmatrix}0 & 1\\ 2 & 1\end{pmatrix}, a_{18} = \begin{pmatrix}0 & 2\\ 1 & 1\end{pmatrix}, a_{19} = \begin{pmatrix}1 & 1\\ 2 & 0\end{pmatrix}, a_{20} = \begin{pmatrix}1 & 2\\ 1 & 0\end{pmatrix},\\ &a_{21} = \begin{pmatrix}2 & 0\\ 1 & 2\end{pmatrix}, a_{22} = \begin{pmatrix}2 & 0\\ 2 & 2\end{pmatrix}, a_{23} = \begin{pmatrix}2 & 1\\ 0 & 2\end{pmatrix}, a_{24} = \begin{pmatrix}2 & 2\\ 0 & 2\end{pmatrix}\Big\}\end{aligned}$$

Dikarenakan himpunan $SL(2,3)$ berhingga, akan lebih mudah untuk mengidentifikasi sifat-sifat yang dimiliki dengan menggunakan Tabel Cayley, yaitu tabel yang akan menampilkan hasil operasi perkalian matriks pada $SL(2,3)$. Hasil perhitungan pada Tabel Cayley tersebut dapat digunakan untuk mengidentifikasi bahwa aksioma grup berlaku pada $SL(2,3)$. Aksioma grup yang dimaksud adalah sifat tertutup, sifat assosiatif, memiliki elemen identitas, dan setiap elemen memiliki invers,

Untuk memudahkan, berikut diberikan beberapa contoh sebagai penjelasan notasi bilangan yang muncul pada Tabel Cayley tersebut. Elemen $a_2 = \begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}$ yang dioperasikan dengan elemen $a_{12} = \begin{pmatrix}0 & 2\\ 1 & 0\end{pmatrix}$ akan memberikan hasil sebagai berikut.

$$a_2 a_{12} = \begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}\begin{pmatrix}0 & 2\\ 1 & 0\end{pmatrix} = \begin{pmatrix}0 & 4\\ 2 & 0\end{pmatrix} \equiv \begin{pmatrix}0 & 1\\ 2 & 0\end{pmatrix} = a_{11},$$

merupakan operasi elemen pada baris a_2 dengan elemen pada kolom a_{12} pada Tabel Cayley. Hasil operasi kedua elemen tersebut adalah elemen a_{11}, untuk menghemat tempat penulisan maka hanya ditulis angka 11. Contoh yang lain, hasil perkalian elemen $a_{15} = \begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix}$ dengan elemen $a_{19} = \begin{pmatrix}1 & 1\\ 2 & 0\end{pmatrix}$ dapat dilihat pada baris ke-15 dan kolom ke-19, yaitu 10, artinya hasil operasi kedua elemen tersebut adalah $a_{10} = \begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix}$. Lebih lengkap, hasil Tabel Cayley ditampilkan pada Tabel 1 berikut ini.

Journal of Data Analysis
Vol.2, No.2, December 2019, p. 53 - 60

Tabel 1 Tabel Cayley Operasi Perkalian Matriks di $SL(2,3)$

	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18	a19	a20	a21	a22	a23	a24
a1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	16	18	19	20	21	22	23	24
a2	2	1	18	17	22	24	21	20	19	23	12	11	16	15	14	13	4	3	9	8	7	5	10	6
a3	3	18	9	14	13	17	23	22	1	11	21	7	20	24	6	8	15	19	2	5	10	16	12	4
a4	4	17	13	8	24	12	14	1	21	18	5	22	23	19	9	10	20	16	7	2	15	6	3	11
a5	5	22	11	18	7	13	1	15	24	20	17	4	19	10	23	9	3	12	6	14	2	21	8	16
a6	6	24	22	13	15	10	17	19	12	1	8	20	21	3	18	7	16	5	11	9	4	14	2	23
a7	7	21	17	12	1	19	5	23	16	14	3	18	6	20	8	24	11	4	13	10	22	2	15	9
a8	8	20	23	1	11	22	19	4	15	16	24	6	3	7	21	18	2	10	14	17	9	12	13	5
a9	9	19	1	24	20	15	12	16	3	21	10	23	5	4	17	22	6	2	18	13	11	8	7	14
a10	10	23	14	21	18	1	16	11	20	6	19	9	4	22	5	17	7	15	8	12	13	3	24	2
a11	11	12	24	10	19	3	8	21	5	17	2	1	14	16	13	15	23	6	22	7	20	9	4	18
a12	12	11	6	23	9	18	20	7	22	4	1	2	15	13	16	14	10	24	5	21	8	19	17	3
a13	13	16	21	19	23	20	3	6	4	5	15	14	2	11	12	1	9	7	17	24	18	10	22	8
a14	14	15	20	22	4	7	24	3	10	19	13	16	12	2	1	11	5	8	23	18	6	17	9	21
a15	15	14	8	5	17	21	6	18	23	9	16	13	11	1	2	12	22	20	10	3	24	4	19	7
a16	16	13	7	9	10	8	18	24	17	22	14	15	1	12	11	2	19	21	4	6	3	23	5	20
a17	17	4	16	20	6	11	15	2	7	3	22	5	10	9	19	23	8	13	21	1	14	24	18	12
a18	18	3	19	15	16	4	10	5	2	12	7	21	8	6	24	20	14	9	1	22	23	13	11	17
a19	19	9	2	6	8	14	11	13	18	7	23	10	22	17	4	5	24	1	3	16	12	20	21	15
a20	20	8	10	2	12	5	9	17	14	13	6	24	18	21	7	3	1	23	15	4	19	11	16	22
a21	21	7	4	11	2	9	22	10	13	15	18	3	24	8	20	6	12	17	16	23	5	1	14	19
a22	22	5	12	3	21	16	2	14	6	8	4	17	9	23	10	19	18	11	24	15	1	7	20	13
a23	23	10	15	7	3	2	13	12	8	24	9	19	17	5	22	4	21	14	20	11	16	18	6	1
a24	24	6	5	16	14	23	4	9	11	2	20	8	7	18	3	21	13	22	12	19	17	15	1	10

Beberapa karakteristik yang dapat dibaca dari Tabel Cayley tersebut adalah hasil perkalian antar dua elemen di $SL(2,3)$ masih merupakan elemen di $SL(2,3)$, artinya $SL(2,3)$ bersifat tertutup terhadap perkalian. Selain itu, dalam tiap baris atau kolom setiap elemen hanya muncul satu kali. Hasil perkalian pada baris pertama (kolom pertama) sama dengan baris (kolom) utamanya, artinya elemen a_1 merupakan elemen identitas di $SL(2,3)$. Dapat juga diamati, bahwa setiap elemen memiliki invers, yaitu dengan melihat kemunculan angka 1 dalam hasil perkaliannya. Dapat diketahui, a_{20} memiliki invers a_{17} karena $a_{20}a_{17} = a_{17}a_{20} = a_1$.

Telah diketahui bahwa o $\big(SL(2,3)\big)$ adalah 24, dan bilangan-bilangan berikut merupakan pembagi bilangan 24, yaitu 1,2,3,4,6,8,12, dan 24. Dengan demikian, berdasarkan Teorema Lagrange, maka kemungkinan order subgrup dari $SL(2,3)$ adalah bilangan-bilangan tersebut.

Dua subgrup trivial dari $SL(2,3)$ adalah $SL(2,3)$ itu sendiri dan $\{a_1\}$, dengan demikian, terdapat subgrup berorder 1 dan 24. Dapat juga langsung dihitung, bahwa $o(a_2) = 2$, dengan demikian terdapat subgrup berorder 2, yaitu $C = \{a_1, a_2\}$. Berdasarkan pengamatan Tabel Cayley, diperoleh juga bahwa salah satu sifat istimewa dari subgrup C tersebut adalah merupakan pusat dari $SL(2,3)$, yaitu elemen di C bersifat komutatif terhadap semua elemen di $SL(2,3)$. Hal tersebut dapat ditulis sebagai $a_1a_j = a_ja_1$ dan $a_2 = a_j = a_ja_2$ untuk semua $a_j \in SL(2,3)$. Dengan demikian, $C = \{a_1, a_2\} = Z\big(SL(2,3)\big)$.

Berdasarkan perhitungan langsung, juga diperoleh bahwa

$$o(a_3) = o(a_4) = o(a_5) = o(a_6) = o(a_7) = o(a_8) = o(a_9) = o(a_{10}) = 3,$$

dengan demikian terdapat setidaknya 8 subgrup berorder 3. Salah satunya, sebagai contoh yang dibangun oleh elemen a_4, yaitu $\langle a_4\rangle = \{a_4, a_4^2 = a_8, a_4^3 = a_1\}$. Selain itu, dapat dihitung secara langsung bahwa

$$o(a_{11}) = o(a_{12}) = o(a_{13}) = o(a_{14}) = o(a_{15}) = o(a_{16}) = 4,$$

artinya $\langle a_{12}\rangle = \{a_{12}, a_{12}^2 = a_2, a_{12}^3 = a_{11}, a_{12}^4 = a_1\}$ merupakan salah satu subgrup berorder 4 di $SL(2,3)$. Subgrup lain yang berorder 4 dibangun oleh salah satu elemen $a_{11}, a_{13}, a_{14}, a_{15}$ atau a_{16}.

Untuk subgrup berorder 6, dapat dihitung bahwa

$$o(a_{17}) = o(a_{18}) = o(a_{19}) = o(a_{20}) = o(a_{21}) = o(a_{22}) = o(a_{23}) = o(a_{24}) = 6.$$

Dengan demikian, akan dapat dibentuk salah satu subgrup berorder 6, yang dibangun oleh a_{22} sebagai berikut

$$\langle a_{22}\rangle = \big\{a_{22}, a_{22}^2 = a_7, a_{22}^3 = a_2, a_{22}^4 = a_5, a_{22}^5 = a_{21}, a_{22}^6 = a_1\big\}.$$

Sementara, sudah diketahui beberapa subgrup berorder $1, 24, 2, 3, 4$, dan 6. Dengan subgrup berorder 1 dan 24 adalah subgrup trivial dan subgrup berorder $2, 3, 4$, dan 6 adalah subgrup siklik. Karena order terbesar elemen $SL(2,3)$ adalah 6 maka tidak ada subgrup siklik yang beroder lebih besar dari 6.

Dengan demikian, selanjutnya subgrup yang mungkin adalah subgrup tidak siklik, yaitu dengan membentuk subgrup yang dibangun lebih dari satu elemen. Caranya dengan membentuk subgrup yang dibangun oleh elemen berorder 2 dan elemen berorder 3. Misal, dengan memilih elemen a_4 sebagai elemen berorder 3 akan diperoleh.

Tabel 2 Subgrup dibangun $\langle a_2, a_4\rangle$

	a1	a2	a4	a8	a17	a20
a1	1	2	4	8	16	20
a2	2	1	17	20	4	8
a4	4	17	8	1	20	2
a8	8	20	1	4	2	17
a17	17	4	20	2	8	1
a20	20	8	2	17	1	4

Diperoleh $\langle a_2, a_4\rangle = \{a_1, a_2, a_4, a_8, a_{17}, a_{20}\}$, namun dapat diperhatikan lebih jauh subgrup tersebut persis sama $\langle a_{17}\rangle = \langle a_{20}\rangle$. Artinya, subgrup yang terbentuk merupakan subgrup siklik. Demikian juga, jika dipilih elemen berorder 3 lain, akan diperoleh

$$\langle a_2, a_3\rangle = \{a_1, a_2, a_3, a_9, a_{18}, a_{19}\} = \langle a_{18}\rangle = \langle a_{19}\rangle$$
$$\langle a_2, a_5\rangle = \{a_1, a_2, a_5, a_7, a_{21}, a_{22}\} = \langle a_{21}\rangle = \langle a_{22}\rangle$$

dan

$$\langle a_2, a_6\rangle = \{a_1, a_2, a_6, a_{10}, a_{23}, a_{24}\} = \langle a_{23}\rangle = \langle a_{24}\rangle.$$

Dengan demikian, $SL(2,3)$ tidak memiliki subgrup tidak siklik berorder 6. Sementara membangun subgrup dengan elemen berorder 2 dan berorder 4 tetap akan diperoleh subgrup siklik beroder 4.

$$\langle a_2, a_{11}\rangle = \{a_1, a_2, a_{11}, a_{12}\} = \langle a_{11}\rangle = \langle a_{12}\rangle$$
$$\langle a_2, a_{13}\rangle = \{a_1, a_2, a_{13}, a_{16}\} = \langle a_{13}\rangle = \langle a_{16}\rangle$$

dan

$$\langle a_2, a_{14}\rangle = \{a_1, a_2, a_{14}, a_{15}\} = \langle a_{14}\rangle = \langle a_{15}\rangle.$$

Demikian juga, subgrup yang dibangun oleh elemen berorder 2 dan elemen berorder 6 merupakan subgrup siklik beroder 6. Sebagai contoh,

$$\langle a_2, a_{17}\rangle = \{a_1, a_2, a_4, a_8, a_{17}, a_{20}\} = \langle a_2, a_4\rangle = \langle a_{17}\rangle = \langle a_{20}\rangle.$$

Selanjutnya, subgrup berorder 8 dapat langsung diamati dari Tabel 1, bahwa perkalian antar elemen berorder 4 menghasilkan elemen berorder 4 dan elemen a_1 dan a_2. Dapat dilihat pada Tabel berikut.

Tabel 3 Tabel Cayley untuk Subgrup Berorder 8

	a1	a2	a11	a12	a13	a14	a15	a16
a1	1	2	11	12	13	14	15	16
a2	2	1	12	11	16	15	14	13
a11	11	12	2	1	14	16	13	15
a12	12	11	1	2	15	13	16	14
a13	13	16	15	14	2	11	12	1
a14	14	15	13	16	12	2	1	11
a15	15	14	16	13	11	1	2	12
a16	16	13	14	15	1	12	11	2

Dengan demikian, diperoleh subgrup berorder 8, yaitu

$$\{a_1, a_2, a_{11}, a_{12}, a_{13}, a_{14}, a_{15}, a_{16}\}.$$

Bagian terakhir dari artikel ini adalah membuktikan bahwa $SL(2,3)$ tidak memiliki subgrup berorder 12. Untuk memudahkan, maka diperlukan beberapa dalil berikut. Dalil pertama terkait grup hingga yang berorder genap.

Dalil 1. *Misalkan G grup hingga berorder genap, maka G memuat paling sedikit satu elemen yang berorder* 2.

Bukti. *Jika $a \in G$ dengan $o(a) = 2$, maka $a^2 = e$ atau $a = a^{-1}$. Tulis G sebagai gabungan yang saling asing sedemikian sehingga*

$$G = \{a \in G | a \neq a^{-1}\} \cup \{e\} \cup \{a \in G | a \neq e, a = a^{-1}\}.$$

Karena order dari $\{a \in G | a \neq a^{-1}\}$ adalah genap dan order dari $\{e\}$ adalah ganjil maka haruslah order dari $\{a \in G | a \neq e, a = a^{-1}\}$ adalah ganjil. Dengan demikian, terdapat paling sedikit satu elemen dari G yang berorder 2. ■

Dalil 2 berikut diadopsi dari soal latihan yang ditulis oleh Shariari dalam buku karangannya [6]
Dalil 2. *Misalkan G adalah grup hingga dan H subgrup G. Asumsikan bahwa banyaknya elemen H adalah setengah dari banyaknya elemen G.*

(a) Jika $x \in G$ maka $x^2 \in H$,
(b) Jika $y \in G$ dengan $o(y) = 3$ maka $y \in H$.

Bukti. *Karena banyaknya koset kiri sama dengan banyaknya koset kanan,maka akan bukti dapat dilakukan dengan salah satu koset saja. Diasumsikan bahwa banyaknya elemen H adalah setengah dari banyaknya elemen G, artinya hanya terdapat dua koset yang saling berbeda dari H di G.*

(a) Ambil sebarang $x \in G$, jika $x \in H$ maka tidak ada yang perlu dibuktikan. Andaikan $x \notin H$ maka dapat ditulis $G = H \cup xH$. Karena $x^2 \in G$ maka haruslah $x^2 \in H$ atau $x^2 \in xH$, jika $x^2 \in H$ tidak ada yang perlu dibuktikan, jika $x^2 \in xH$ berakibat $x^2 = xh$ untuk suatu $h \in H$, artinya $x = h \in H$, kontradiksi dengan $x \notin H$. Artinya, untuk setiap $x \in G$ maka $x^2 \in H$.

(b) Ambil sebarang $y \in G$ dengan $o(y) = 3$. Andaikan $y \notin H$, maka dapat ditulis $G = H \cup yH$. Karena $y^2 \in G$ maka haruslah $y^2 \in H$ atau $y^2 \in yH$, jika $y^2 \in H$ maka

$$y = ey = y^3y = y^4 = (y^2)^2 \in H$$

Kontradiksi dengan $y \in H$. Jika $y^2 \in yH$ maka $y^2 = yh$ untuk suatu $h \in H$, diperoleh $y \in H$, juga kontradiksi dengan $y \in H$. Artinya, jika $y \in G$ dengan $o(y) = 3$ maka $y \in H$. ■

Selanjutnya akan dibuktikan bahwa $SL(2,3)$ tidak memiliki subgrup berorder 12. *Bukti yang diberikan merupakan detail dari langkah-langkah yang ditulis oleh Shariari* [6].
Dalil 3. *Grup $SL(2,3)$ tidak memiliki subgrup berorder* 12.

Bukti. *Langkah pertama adalah menghitung banyaknya elemen $SL(2,3)$, yang telah dilakukan pada bagian awal hasil pembahasan, bahwa order dari $SL(2,3)$ adalah 24. Langkah kedua adalah menghitung banyak elemen di $SL(2,3)$ yang berorder 3, telah diperoleh hasil bahwa terdapat delapan elemen yang berorder 3, yaitu $a_3, a_4, a_5, a_6, a_7, a_8, a_9$, dan a_{10}. Langkah ketiga adalah menemukan elemen berorder 2 dan elemen tersebut terdapat dalam center dari $SL(2,3)$. Langkah keempat, terakhir, misalkan terdapat subgrup H berorder 12 di $SL(2,3)$, maka berdasarkan Dalil 1, subgrup H akan memuat elemen berorder 2, dengan demikian, elemen $a_2 \in H$. Dan berdasarkan Dalil 2, semua elemen berorder 3 juga termuat di H. Padahal, untuk menjamin sifat tertutup di H maka haruslah semua elemen berorder 6, yaitu hasil kali elemen a_2 dengan sebarang elemen berorder 3, juga termuat di H. Hal ini berakibat, elemen-elemen berikut harus termuat di H, yaitu*

$$a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{17}, a_{18}, a_{19}, a_{20}, a_{21}, a_{22}, a_{23}, a_{24}.$$

Satu elemen identitas, satu elemen berorder 2, delapan elemen berorder 3, dan delapan elemen berorder 6, artinya subgrup H setidaknya memuat $1 + 1 + 8 + 8 = 18$. Jadi, tidak mungkin terdapat subgrup H yang berorder 12 di $SL(2,3)$. ■

Dengan demikian, telah dibuktikan bahwa konvers Teorema Lagrange tidak berlaku. Artinya, meskipun suatu order grup memiliki pembagi bilangan k, belum tentu grup tersebut memiliki subgrup yang berorder k. Dalam kasus $SL(2,3)$, yang merupakan grup berorder 24, dan bilangan 12 merupakan pembagi bilangan 24, tetapi $SL(2,3)$ tidak memiliki subgrup yang berorder 12.

4. Kesimpulan

Berdasarkan hasil dan pembahasan, dapat disimpulkan bahwa grup $SL(2,3)$ memiliki 24 elemen dan merupakan grup tidak komutatif. Grup $SL(2,3)$ memiliki beberapa subgrup berorder $1, 2, 3, 4, 6, 8$ dan 24. Subgrup berorder 1 dan 24 merupakan subgrup trivial, sementara subgrup

tak trivial adalah subgrup berorder $2, 3, 4, 6$ yang merupakan subgrup siklik, dan subgrup berorder 8 bukan merupakan subgrup siklik.

Meskipun bilangan 12 merupakan pembagi bilangan 24, order dari grup $SL(2,3)$, grup tersebut tidak memiliki subgrup berorder 12. Dengan demikian, grup $SL(2,3)$ merupakan salah satu contoh bahwa konvers Teorema Lagrange tidak berlaku. Sebagai saran, penelitian lebih lanjut dapat dilakukan pada grup $SL(2,5)$.

Daftar Pustaka

[1] J. A. Gallian, *Contemporary Abstract Algebra*, 9th ed. Boston: Cengage Learning, 2017.
[2] M. Brennan and D. Machale, "Variations on a Theme: A_4 Definitely Has no Subgroup of Order Six!," *Math. Mag.*, vol. 73, no. 1, p. 36, Feb. 2000.
[3] J. B. Nganou, "How rare are subgroups of index 2?," *Math. Mag.*, vol. 85, no. 3, pp. 215–220, Jun. 2012.
[4] G. Mackiw, "Finite Groups of 2×2 Integer Matrices," *Math. Mag.*, vol. 69, no. 5, pp. 356–361, 1996.
[5] G. Mackiw, "The Linear Group SL (2, 3) as a Source of Examples," *Math. Gaz.*, vol. 81, no. 490, p. 64, Mar. 1997.
[6] Shariar Shariari, *Algebra in Action, A Course in Groups, Rings, and Fields*. Providence, Rhode Island: American Mathematical Society, 2017.
[7] L. Gilbert and J. Gilbert, *Elements of Modern Algebra.*, 8th ed. Stamford: Cengage Learning, 2015.
[8] T. W. Hungerford, *Abstract Algebra, an Introduction*, Third. Boston: Books/Cole, Cengage Learning, 2014.
[9] M. Mahmudi, "Suatu Catatan tentang Subgrup Normal dan Ideal," *J. Data Analysis*, vol. 1, no. 2, pp. 77–82, Dec. 2018.

Journal of Data Analysis
Vol.2, No.2, December 2019, p. 61-70

Penerapan *Time Delay Neural Network* pada Model Akustik untuk Sistem *Voice-to-Text* Berbahasa Sunda

Alim Misbullah[1*], Nazaruddin[1], Marzuki[2] dan Zulfan[1]

[1] Jurusan Informatika, Fakultas MIPA, Universitas Syiah, Banda Aceh, Indonesia
[2]Jurusan Statistika, Fakultas MIPA, Universitas Syiah Kuala, Banda Aceh, Indonesia
E-mail: misbullah@unsyiah.ac.id*, anzaro@unsyiah.ac.id, marzuki@unsyiah.ac.id, zulfan.abdullah@unsyiah.ac.id

* = corresponding author

Abstrak

Penerapan metode *deep learning* dalam berbagai bidang terutama pada kasus pengenalan pola sudah menghasilkan akurasi yang sangat menjanjikan. Jaringan saraf tiruan atau *neural network* merupakan bagian dari *deep learning* yang digunakan untuk melatih model pada kasus pengenalan pola seperti model untuk sistem pengenalan ucapan (*voice-to-text*). *Neural network* akan menyimpan informasi dari setiap fitur data berupa bobot pada jaringan yang terhubung antar-*layer* pada model yang dibangun. Bobot pada jaringan tersebut diperbaharui berdasarkan banyaknya fitur dari data yang diinput. Sistem *voice-to-text* merupakan salah satu bidang pengenalan pola yang mengimplementasikan *neural network* untuk membangun model akustik. Model akustik pada sistem pengenalan ucapan dilatih menggunakan data audio berupa percakapan atau rekaman dari setiap individu untuk bahasa tertentu seperti bahasa Inggris. Penerapan *neural network* untuk sistem pengenalan ucapan berbahasa Inggris sudah banyak dilakukan bahkan sudah diimplementasikan dalam bentuk aplikasi karena mampu menghasilkan akurasi yang tinggi. Namun, penggunaan *neural network* untuk bahasa lokal masih jarang digunakan. Dalam tulisan ini, *time delay neural network* digunakan untuk membangun model akustik pada sistem pengenalan ucapan berbahasa Sunda. Berdasarkan hasil pengujian terhadap model akustik, *time delay neural network* mampu menghasilkan *WER* sampai dengan 0.57% setelah dilakukan penyesuaian pada *hyperparameter* dari *neural network*.

Informasi Artikel

Sejarah Artikel:
Diajukan, 19 Des 2019
Diterima, 2 Apr 2020

Kata Kunci:
Deep Learning,
Neural Network,
Model Akustik,
Sistem Pengenalan Ucapan

Abstract

Implementation of deep learning techniques has given promising results recently in any research area, especially for pattern recognition. Neural network as a part of deep learning has been widely used to build model for various pattern recognition field including speech recognition. In neural network, weights which is parameters among layers play important roles to capture information from input data. The parameters are updated frequently based on input features in each iteration. In speech recognition, neural network is implemented to build acoustic model that uses speech from different speakers as training data. The acoustic model is built for specific language such as English, Mandarin and Indonesian. In recent years, the speech recognition

Keyword:
Deep Learning
Neural Network
Acoustic Model
Speech Recognition System

system using deep neural network for English language has been developed well and use in many applications. But, implementation of deep neural network for local language is rarely done. In this research, time delay neural network is used to build acoustic model for speech recognition system of Sundanese language. Based on experimental result, the implementation of time delay neural network can reduce WER to be 0.57% with well-tuned hyperparameters of neural network.

1. Pendahuluan

Jaringan saraf tiruan (*neural network*) saat ini telah memberikan hasil yang menjanjikan pada berbagai bidang teknologi terutama untuk kasus pengenalan pola seperti sistem pengenalan ucapan (*voice-to-text*). Umumnya, struktur *neural network* terdiri dari *input layer*, *hidden layer* dan *output layer*. Setiap *node* pada *input layer* merepresentasikan vector untuk jumlah data input, sedangkan setiap *node* pada *hidden layer* berfungsi untuk mengontrol dapat atau tidaknya informasi dari *input layer* diteruskan ke *layer* berikutnya. Pada *output layer*, setiap *node* didefinisikan sebagai target dari kelas yang akan diprediksi.

Sistem *voice-to-text* memiliki 2 (dua) komponen utama untuk membangun sistem yaitu model akustik (*acoustic model*) dan model bahasa (*language model*). Model akustik merupakan sebuah model yang mengandung representasi statistik dari setiap *voice* yang berbeda dalam membentuk sebuah kata atau kalimat. Setiap representasi statistik tersebut menentukan sebuah label atau target yang disebut dengan suku kata (*phonemes*).

Struktur *neural network* pada model akustik digunakan untuk melatih model sehingga mampu menyimpan informasi dari setiap fitur data *voice* yang diinput. Struktur *neural network* menggunakan pendekatan *learning based* dalam melatih model akustik yang berarti bahwa setiap fitur akan melalui setiap *hidden layer* sehingga informasi yang ada di dalam fitur akan tercatat pada setiap jaringan yang terhubung antar-*layer* yang disebut dengan parameter (*weight*). Selanjutnya, parameter tersebut akan diperbaharui berdasarkan *error* yang didapat antara nilai keluaran (*output*) dan nilai target *(class)*.

Proses melatih model akustik menggunakan *neural network* memiliki 2 (dua) tahapan utama yaitu *feedforward* dan *backpropagation*. Pada tahapan *feedforward*, setiap data input berupa vektor akan melalui setiap *hidden layer* dalam *neural network* sampai dengan *output layer*. Setiap *node* pada *hidden layer* akan menjadi aktif atau tidak berdasarkan transformasi hasil perkalian linier antara vector input dan parameter matriks yang ada pada jaringan. Proses transformasi hasil perkalian tersebut menggunakan sebuah fungsi yang disebut sebagai fungsi aktifasi (*activation function*). Sedangkan pada tahap *backpropagation*, parameter matriks akan diperbaharui berdasarkan selisih antara nilai *output* dan nilai target yang ada yang disebut dengan nilai *error*. Nilai error ini menjadi acuan awal untuk menentukan nilai parameter baru pada setiap jaringan.

Penggunaan *neural network* untuk membangun model pada sistem *voice-to-text* sudah banyak ditemukan terutama untuk sistem *voice-to-text* berbahasa Inggris [1, 2, 3, 4]. Bahkan sistem *voice-to-text* berbahasa Inggris sudah banyak digunakan pada aplikasi dalam kehidupan seperti Google Asisten, Microfost Cortana, dan Alexa Amazon. Sedangkan, penelitian sistem *voice-to-text* untuk bahasa lokal seperti bahasa Sunda masih jarang dilakukan karena keterbatasan data *voice* yang tersedia dan bahkan akurasinya masih perlu ditingkatkan.

Uraian di atas mengantarkan peneliti untuk mengimplementasikan *neural network* untuk membangun model akustik pada sistem *voice-to-text* berbahasa Sunda sehingga dapat diperoleh akurasi yang lebih baik dari hasil penelitian yang telah ada sebelumnya. Hasil penelitian ini dapat digunakan sebagai *prototipe* awal dalam membangun aplikasi *voice-to-text* berbahasa Sunda untuk berbagai perangkat.

2. Tinjauan Kepustakaan

Penelitian-penelitian *voice-to-text* untuk bahasa Sunda sudah pernah dilakukan sebelumnya. Sakti dkk [14] mengembangkan pengenalan ucapan berbasis *Grapheme* terhadap bahasa Jawa, Sunda, Bali, dan Batak. Tujuan akhir dari penelitian-penelitian tersebut adalah ingin membuat alat penerjemah ucapan dari bahasa daerah ke bahasa Inggris dan bahasa Indonesia. Hasil penelitian menunjukkan bahwa dengan menggunakan *multilingual acoustic model*, performa dapat meningkatkan pada aksen bahasa Indonesia, tetapi tidak untuk bahasa daerah. Ini menunjukkan bahwa karakteristik akustik bahasa Indonesia cukup berbeda dari karakteristik akustik dari bahasa-bahasa daerah tersebut.

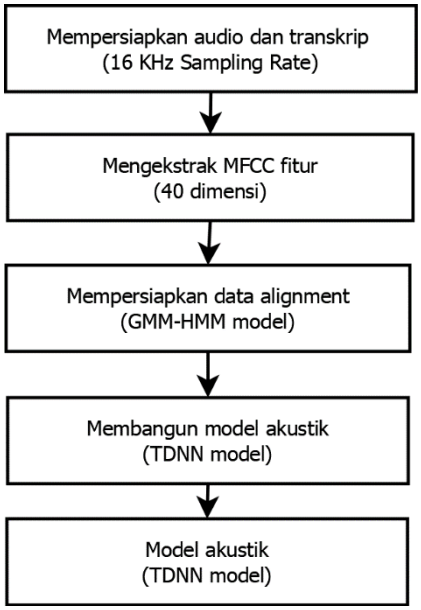
Rahmawati dkk [14] melakukan penelitian yang bertujuan untuk membandingkan fitur dan teknik pemodelan terbaik untuk mengenali dialek bahasa Jawa dan Sunda pada bahasa Indonesia menggunakan MFCC dan kombinasi dari MFCC + pitch. Selain itu juga ingin membandingkan hasil teknik pemodelan dengan menggunakan GMM dan *i-vector*. Berdasarkan hasil penelitian tersebut dapat disimpulkan bahwa fitur terbaik untuk mengenali dialek bahasa Jawa dan Sunda dari ucapan bahasa Indonesia adalah kombinasi dari fitur MFCC + *pitch* dengan teknik pemodelan terbaik adalah *i-vector*.

Teknik *Deep Neural Network* juga pernah diuji coba untuk pengenalan ucapan bahasa Sunda dialek Utara [15], bahasa Sunda dialek Tengah Timur [16] dan bahasa Sunda dialek Garut [17]. Namun dataset yang mereka uji jumlahnya sangat kecil sehingga tidak dapat ditarik kesimpulan yang pasti dari kinerja *Deep Neural Network* terhadap bahasa Sunda.

Kjartansson dkk [18] yang tergabung dalam *Google Research* telah menyediakan dataset untuk 5 (lima) bahasa daerah yaitu bahasa Jawa, Sunda, Sinhala, Nepal, dan Bengali Bangladesh. Setiap bahasa terdiri dari rata-rata sekitar 200.000 ucapan yang direkam oleh sukarelawan penutur asli *(native*) di wilayah masing-masing. Dataset tersebut diklaim sangat cocok untuk penelitian pemodelan akustik pada sistem pengenalan ucapan. Dataset tersebut diizinkan untuk digunakan oleh para peneliti untuk mengembangkan sistem pengenalan ucapan untuk bahasa-bahasa tersebut. Akurasi untuk bahasa Sunda dalam penelitian ini masih sebatas pada model GMM-HMM saja sedangkan untuk melatih model akustik, tidak dilanjutkan menggunakan *Deep Neural Netwok* sehingga performa model masih sangat mungkin untuk ditingkatkan. Model akustik dibangun menggunakan *Time Delay Deep Neural Networks* untuk dataset berbahasa Sunda [18]. Struktur *Time Delay Neural Network* diuji pada jumlah *node* berbeda dari setiap *layer*.

3. Metode Penelitian

Penelitian ini dibagi ke dalam 2 (dua) tahapan yaitu pelatihan dan pengujian model. Tahapan pertama adalah melatih model akustik untuk sistem *voice-to-text* berbahasa Sunda menggunakan *time delay neural network* disingkat dengan TDNN [4]. Model akustik akan dibangun menggunakan *framework* untuk sistem *voice-to-text* yaitu Kaldi toolkit [6]. Proses membangun model akustik dimulai dengan menyiapkan data *voice* dan transkrip, mengekstrak fitur, melakukan *alignment* dengan model GMM-HMM [5], dan membangun model dengan TDNN. Gambar 1 menunjukkan tahapan pertama yang dilakukan untuk membangun model akustik dalam penelitian ini.



Gambar 1 Tahapan membangun model akustik

Pembangunan model akustik untuk sistem *voice-to-text* terdiri dari beberapa tahapan dengan penjelasan sebagai berikut:

1. Mempersiapkan audio dan transkrip
   Data audio dan transkrip dipersiapkan dalam format *wav* dengan *sampling rate* 16000 Hz dan memiliki *mono channel*. Selanjutnya, dataset tersebut disesuaikan formatnya dengan Kaldi *toolkit* sehingga akan mudah dalam proses di tahapan selanjutnya.
2. Mengekstrak MFCC fitur
   Umumnya, fitur yang digunakan untuk melatih model akustik pada sistem *voice-to-text* adalah MFCC (*Mel-Frequency Cepstral Coefficient*) [7]. Fitur MFCC tersebut mudah digunakan dan mampu meningkatkan akurasi untuk *voice based* sistem [8, 9].
3. Mempersiapkan data alignment
   Data alignment adalah proses pemetaan *voice* dari setiap *frame* direntan waktu tertentu terhadap suku katanya. Proses tersebut dilakukan menggunakan GMM-HMM (*Gaussian Mixture Model – Hidden Markov* Model) model [10] yang juga dibangun menggunakan data *voice* yang memiliki transkrip.
4. Membangun model akustik
   Model akustik dapat dibangun menggunakan GMM-HMM model, namun akurasi yang dihasilkan tidak sebaik model yang dibangun menggunakan *neural network*. Pada tahapan ini, model akustik akan dibangun menggunakan *time delay neural network* (TDNN) [4]. Suku kata yang diperoleh pada tahapan sebelumnya akan menjadi target untuk setiap input fitur yang dimasukkan ke dalam struktur TDNN.

3.1. Persiapan Voice Dataset

Model akustik dan model bahasa dibangun menggunakan data *voice* berbahasa Sunda[1]. Data tersebut merupakan bagian dari projek Google Inc. yang dikumpulkan menggunakan *tool* yang disebut *DataHound* [12]. Tabel 1 menunjukkan informasi mengenai data *voice* berbahasa Sunda yang digunakan dalam penelitian ini.

[1] https://openslr.org/36/

Tabel 1 Distribusi data *voice* berbahasa Sunda

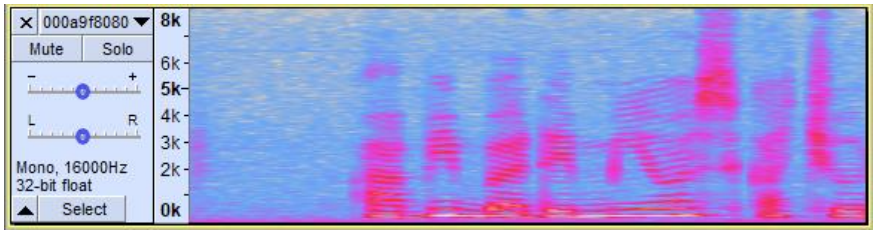
Dataset	# *Speaker*	# *Utterances*	Total Waktu (*hours*)
Training	537	216112	327
Tessting	6	1994	3,2
Total	543	218106	329,2

Selanjutnya, data *voice* tersebut dikonversi ke dalam format yang dapat digunakan oleh Kaldi. Format yang digunakan oleh Kaldi memuat informasi yang terdiri dari *text*, *wav.scp*, *utt2spk*, dan *spk2utt*. Isi dari setiap *file* yang diperlukan oleh Kaldi untuk membangun model akustik ditunjukkan di dalam Tabel 2. Setiap *file* terdiri dari 2 (dua) kolom yang kolom pertamanya mendefiniskan *utterance_id* untuk setiap ucapan. Bagian kolom pertama tersebut akan membentuk relasi antar *file* yang berikutnya digunakan untuk membangun model akustik.

Tabel 2 Format file untuk Kaldi *toolkit*

Format File	Contoh
text	`01f12-0cda366937 wartawan keur nyiar berita di situ cileunca` `01f12-20cbb934dd yana julio olohok ningali marc anthony keur diwawancara`
wav.scp	`01f12-0cda366937 flac -cds /asr_sundanese/data/0c/0cda366937.flac |` `01f12-20cbb934dd flac -cds /asr_sundanese/data/20/20cbb934dd.flac |`
utt2spk	`01f12-0cda366937 01f12` `01f12-20cbb934dd 01f12`
spk2utt	`01f12 01f12-0cda366937 01f12-20cbb934dd`

Setiap file *audio* disimpan dalam format *flac* sehingga memerlukan *library* khusus untuk membacanya. File *audio* tersebut memiliki *sampling rate* 16000 kHz dan *mono channel*. Gambar 2 menunjukkan *spectrogram* untuk salah satu data *audio* berbahasa Sunda. Bagian berwarna merah pada Gambar 1 merepresentasikan energi yang dihasilkan dari ucapan dan energi tersebut mengandung kata pada rentang waktu tertentu.



Gambar 2 Contoh *spectrogram* data *audio* berbahasa Sunda

3.2 Pembuatan Kamus dan Leksikal

Proses membangun model akustik merupakan bagian dari proses klasifikasi sehingga memerlukan label atau kelas yang akan menjadi target dari setiap *frame* pada audio. Label tersebut dapat dibangun menggunakan kamus kata yaitu jumlah kata unik yang terdapat di dalam transkrip dari *training* dan *testing* data. Jumlah kata unik untuk kamus kata berbahasa Sunda dalam penelitian ini adalah sebanyak 12.067 yang diekstrak dari transkrip *training* dan *testing* data. Selanjutnya, kata unik tersebut akan diurai menjadi suku kata (*phonemes*) atau leksikal yang merupakan cara baca dari setiap kata. Tabel 3 menunjukkan contoh kata dan suku kata yang

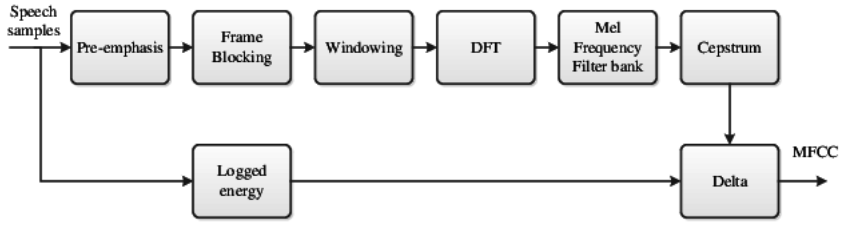
dibangun dari kamus kata. Aturan yang digunakan untuk membangun leksikal adalah berdasarkan huruf vocal dan konsonan yang ada pada sebuah kata, seperti *sayang* akan menjadi /sa/, /ya/, /ng/. Selanjutnya, setiap suku kata akan menjadi label atau kelas untuk proses klasifikasi dalam membangun model akustik. Jumlah suku kata yang berhasil diesktrak adalah sebanyak 276 yang berarti bahwa jumlah kelas untuk proses klasifikasi adalah sebanyak suku kata yang ada.

Tabel 3 Kata dan suku kata berbahasa Sunda

Kata	Suku Kata
adang	a da ng
kadek	ka de k
kakurung	ka ku ru ng
kamari	ka ma ri
trapani	t ra pa ni
teuas	te u a s

3.2. Ekstraksi Fitur

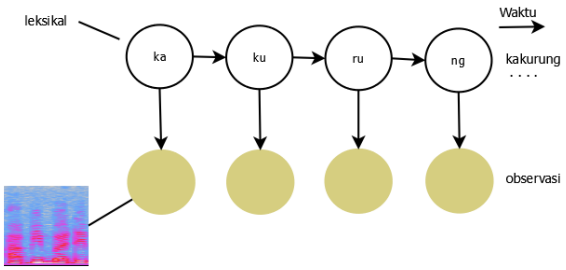
Tahapan ekstraksi fitur merupakan tahapan penting dalam proses membangun model akustik. Fitur yang digunakan untuk membangun model akustik dari data *voice* adalah MFCC (*Mel-Frequency Cepstral Coefficient*). Fitur MFCC akan diekstrak menggunakan *sliding window* sebesar 25 ms (*milliseconds*) dan *overlapping* sebesar 10 ms untuk total waktu dari sebuah data *voice*. Fitur tersebut menggunakan 40 dimensi untuk setiap *frame* yang diekstrak dan disimpan dalam bentuk vector untuk setiap frame. Gambar 3 mengilustrasikan proses yang dilalui untuk memperoleh fitur MFCC dari sebuah data *voice*.



Gambar 3 Ilustrasi ekstraksi fitur MFCC (Sumber [13])

3.3 Data Alignment

Data *alignment* merupakan tahapan untuk mendapatkan target dari setiap *frame* yang telah diekstrak menjadi fitur. Di tahapan ini, model GMM-HMM digunakan untuk mencocokkan setiap *frame* dengan suku kata yang merepresentasikannya. Pada sistem *voice-to-text*, GMM akan memodelkan distribusi fitur untuk setiap suku kata yang ada dengan menghitung nilai *likelihood* untuk setiap distribusi *gaussian*. Setiap suku kata dapat dibentuk oleh beberapa *frame* karena adanya *overlapping* pada tahapan esktraksi fitur. Sedangkan HMM merupakan rantai *markov* yang memuat variabel tertutup antara fitur *voice* dengan suku katanya seperti yang ditunjukkan pada Gambar 4.



Gambar 4 Ilustrasi HMM untuk observasi terhadap leksikalnya

3.4 Pelatihan Model Akustik

Model akustik merupakan sebuah model statistik yang dibangun untuk mendapatkan korelasi antara *frames* dan suku katanya. Secara umum, model-model yang digunakan pada sistem *voice-to-text* dapat direpresentasikan dalam bentuk fungsi distribusi berikut

$$W^* = \arg\max_W P(W|X) \quad (1)$$

$$W^* = \arg\max_W \frac{P(X|W)P(W)}{P(X)} \quad (2)$$

$$W^* = \arg\max_W P(X|W)P(W) \quad (3)$$

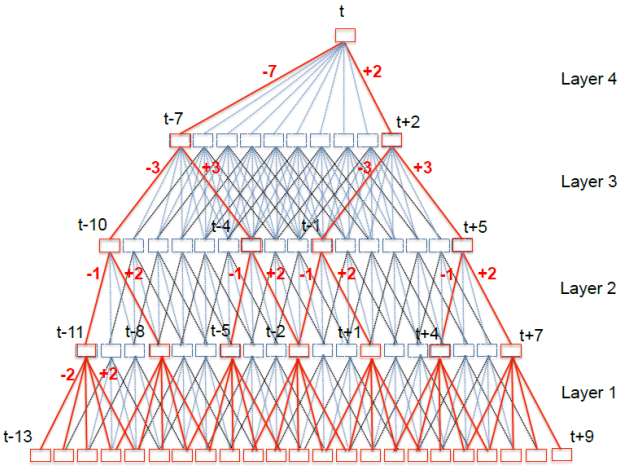
dimana W^*merupakan urutan kata. Fungsi $P(X|W)$ dan $P(W)$ secara berturut-turut adalah model akustik dan model bahasa. Tujuan umum dari fungsi (3) adalah menemukan urutan kata yang sesuai dengan memaksimalkan keluaran dari model akustik dan model bahasa terhadap kata tersebut. Umumnya, fungsi $P(X|W)$ atau model akustik dapat dibangun menggunakan pendekatan GMM-HMM [10], namun model tersebut tidak mampu bekerja optimal untuk data *audio* yang banyak dan *speaker* berbeda-beda [1]. Sedangkan fungsi $P(W)$ adalah model bahasa yang digunakan untuk mendapatkan nilai probabilitas terbaik di antara urutan kata. Gambar 5 mengilustrasikan bagaimana penggunaan model bahasa.

ketua jurusan melakukan kunjungan ke daerah (1)

ketua jurusan kunjungan melakukan ke daerah (2)

Gambar 5 Penggunaan model bahasa untuk urutan kata

Model GMM-HMM pada penelitian ini hanya digunakan untuk melakukan *alignment* dari data *voice*. Selanjutnya, data *alignment* tersebut digunakan untuk melatih model akustik menggunakan *neural network*. Jenis *neural network* yang digunakan untuk melatih akustis adalah TDNN (*Time Delay Neural Networks*) [4]. Pada struktur standar *deep neural network*, waktu permulaan dan akhir dari sebuah *utterance* harus ditentukan sebelum dilakukan proses perkalian linier. Data fitur akan diinput secara keseluruhan ke dalam struktur *deep neural network* sekaligus dalam satu waktu. Sedangkan pada stuktur TDNN, data fitur tidak diinput secara keseluruhan melainkan dibagi ke dalam beberapa bagian pada waktu yang berbeda. Gambar 6 menunjukkan struktur TDNN yang dibangun berdasarkan waktu berbeda untuk setiap *layer*.



Gambar 6 Struktur *Time Delay Neural Network* (Sumber [4])

Selanjutnya, fungsi $P(X|W)$ yang dimodelkan menggunakan GMM-HMM diganti menggunakan TDNN. Secara umum, formulasi untuk setiap *layer* pada *neural network* didefinisikan sebagai berikut.

$$\mathbf{a} = \mathbf{W}_{\boldsymbol{a}} \cdot \mathbf{x} + \mathbf{b} \quad (4)$$

$$\mathbf{z} = relu(\mathbf{a}) \quad (5)$$

$$\mathbf{y} = \text{softmax}(\mathbf{W}_{\boldsymbol{z}} \cdot \mathbf{z} + \mathbf{b}) \quad (6)$$

dimana **a** merupakan hasil perkalian linier antara vektor input **x** dan parameter matrik $\mathbf{W}_{\boldsymbol{a}}$ serta dijumlahkan dengan bias **b**. Sedangkan nilai **z** diperoleh setelah dilakukan aktivasi terhadap nilai **a** menggunakan *relu* sebagai fungsi aktivasi. Nilai **y** merupakan nilai keluaran akhir yang diperoleh dari hasil aktivasi proses perkalian linier antara vektor **z** dan parameter matriks $\mathbf{W}_{\boldsymbol{z}}$ menggunakan fungsi *softmax*. Nilai **y** tersebut kemudian digunakan untuk menghitung selisih terhadap target kelas yang didefiniskan sebelumnya menggunakan data *alignment*. Nilai selisih dapat dihitung menggunakan formulasi berikut.

$$Cross\ Entropy\ Loss = -(y_i \log(\hat{y}_i)) + (1 - y_i)\log(1 - \hat{y}_i) \quad (7)$$

dimana y_i dan $\hat{y}_i$ merupakan nilai prediksi dari *neural network* dan target yang didefinisikan secara berturut-turut.

Tahapan pengujian merupakan tahapan untuk menguji model yang telah dilatih menggunakan data testing. Model akustik yang telah dilatih akan diuji menggunakan data testing sebanyak 1994 ucapan dari 6 *speakers* berbeda. Pada proses pengujian model akustik, mode bahasa juga diperlukan untuk memprediksi urutan kata yang tepat. Model bahasa yang digunakan pada penelitian ini adalah model yang dibangun menggunakan pendekatan *n-gram* [21].

4. Analisis Pembahasan

Model akustik dalam penelitian ini dibangun menggunakan Kaldi *toolkit* [6]. Proses pelatihan model dimulai dari mengekstraksi fitur MFCC kemudian melatih model GMM-HMM untuk memperoleh data *alignment*. Model *i-vector* [19] juga dibangun untuk menyimpan informasi dari setiap *speaker* sehingga kinerja model akustik menjadi lebih baik. Selanjutnya, model akustik dari struktur TDNN dilatih menggunakan *learning rate* sebesar 0.015 dan *epochs* sebesar 3. Tabel 4 menunjukkan WER (*Word Error Rate*) untuk model akustik menggunakan GMM-HMM dan TDNN.

Tabel 4 Performa model akustik

Model Akustik	WER (%)
Monophone	43,58
Triphone	7,19
TDNN – 13L128N	0,59

** WER adalah persentase jumlah kata error yang diprediksi oleh model*

Hasil pengujian pada Tabel 4 menunjukkan bahwa model akustik yang dilatih menggunakan struktur TDNN – 13 *hidden layer* dan 128 *hidden nodes* dapat menurunkan WER mutlak sebesar 6.6%. Pengujian dilanjutkan dengan mengubah *hidden nodes* menjadi lebih besar dan menambahkan *layer SpecAugment* [20] untuk mendapatkan hasil yang lebih baik. Hasil pengujian untuk struktur TDNN berbeda ditunjukkan oleh Tabel 5.

Tabel 5 Performa model akustik menggunakan struktur TDNN

Model Akustik	WER (%)
TDNN – 13L128N	0,59
TDNN – 13L128N + SpecAug	0,57
TDNN – 13L256N + SpecAug	0,51

Penurunan WER mutlak sebesar 0.02% diperoleh dari TDNN – 13 *hidden layers* 128 *hidden nodes* dan dikombinasikan dengan *layer SpecAugment*. Sedangkan hasil terbaik dari model akustik menggunakan TDNN diperoleh dari struktur TDNN dengan 13 *hidden layers* 256 *hidden nodes* dikombinasikan dengan *layer SpecAugment*.

5. Kesimpulan dan Saran

Penelitian ini dapat disimpulkan bahwa penggunaan *Time Delay Neural Network* pada sistem *voice-to-text* dapat meningkatkan kinerja akustik model dibandingkan dengan model GMM-HMM biasa. Penambahan jumlah *hidden nodes* pada struktur TDNN dapat meningkatkan performa model akustik mejadi lebih baik. Penelitian selanjutnya akan dilakukan penambahan data training dengan teknik data augmentasi sehingga model akustik menjadi lebih baik. Selain itu, penggunaan struktur *neural network* yang lain seperti CNN (*Convolutiinal Neural Network*) dan LSTM (*Long Short-Term Memory*) juga akan dilakukan di penelitian selanjutnya.

Daftar Pustaka

[1] Deng, L., Hinton, G., & Kingsbury, B. (2013, May). New types of deep neural network learning for speech recognition and related applications: An overview. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 8599-8603). IEEE.

[2] Vanhoucke, V., Devin, M. and Heigold, G., 2013, May. Multiframe deep neural networks for acoustic modeling. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 7582-7585). IEEE.

[3] Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Kingsbury, B. and Sainath, T., 2012. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal processing magazine*, *29*.

[4] Peddinti, V., Povey, D. and Khudanpur, S., 2015. A time delay neural network architecture for efficient modeling of long temporal contexts. In *Sixteenth Annual Conference of the International Speech Communication Association*.

[5] Seltzer, M.L., Yu, D. and Wang, Y., 2013, May. An investigation of deep neural networks for noise robust speech recognition. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 7398-7402). IEEE.

[6] Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P. and Silovsky, J., 2011. The Kaldi speech recognition

toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding* (No. CONF). IEEE Signal Processing Society.

[7] Ittichaichareon, C., Suksri, S. and Yingthawornsuk, T., 2012, July. Speech recognition using MFCC. In *International Conference on Computer Graphics, Simulation and Modeling* (pp. 135-138).

[8] Dimitriadis, D., Maragos, P. and Potamianos, A., 2005. Robust AM-FM features for speech recognition. *IEEE signal processing letters*, *12*(9), pp.621-624.

[9] Tiwari, V., 2010. MFCC and its applications in speaker recognition. *International journal on emerging technologies*, *1*(1), pp.19-22.

[10] Su, D., Wu, X. and Xu, L., 2010, March. GMM-HMM acoustic model training by a two level procedure with Gaussian components determined by automatic model selection. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 4890-4893). IEEE.

[11] Kjartansson, O., Sarin, S., Pipatsrisawat, K., Jansche, M. and Ha, L., 2018. Crowd-Sourced Speech Corpora for Javanese, Sundanese, Sinhala, Nepali, and Bangladeshi Bengali.

[12] Hughes, T., Nakajima, K., Ha, L., Vasu, A., Moreno, P.J. and LeBeau, M., 2010. Building transcribed speech corpora quickly and cheaply for many languages. In *Eleventh Annual Conference of the International Speech Communication Association.*

[13] Tran, V.H., Nguyen, L.T.T., Hoang, T. and Tran, X.T., 2013. Design and Implementation of a SoPC System for Speech Recognition.

[14] Sakti, S. and Nakamura, S., 2014. Recent progress in developing grapheme-based speech recognition for Indonesian ethnic languages: Javanese, Sundanese, Balinese and Bataks. In *Spoken Language Technologies for Under-Resourced Languages*.

[15] Rahmawati, R. and Lestari, D.P., 2017, October. Java and Sunda dialect recognition from Indonesian speech using GMM and I-Vector. In *2017 11th International Conference on Telecommunication Systems Services and Applications (TSSA)* (pp. 1-5). IEEE.

[16] Arwandani, G., Osmond, A.B. and Nugrahaeni, R.A., 2018. Deep Neural Network Untuk Pengenalan Ucapan Pada Bahasa Sunda Dialek Utara. *eProceedings of Engineering*, *5*(3).

[17] Fathurrahman, D.N., Osmond, A.B. and Saputra, R.E., 2018. Deep Neural Network Untuk Pengenalan Ucapan Pada Bahasa Sunda Dialek Tengah Timur (majalengka). *eProceedings of Engineering*, *5*(3).

[18] Hakim, L.A., Osmond, A.B. and Saputra, R.E., 2018. Recurrent Neural Network Untuk Pengenalan Ucapan Pada Bahasa Sunda Selatan Dialek Garut. *eProceedings of Engineering*, *5*(3).

[19] Kjartansson, O., Sarin, S., Pipatsrisawat, K., Jansche, M. and Ha, L., 2018. Crowd-Sourced Speech Corpora for Javanese, Sundanese, Sinhala, Nepali, and Bangladeshi Bengali.

[20] Senior, A. and Lopez-Moreno, I., 2014, May. Improving DNN speaker independence with i-vector inputs. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 225-229). IEEE.

[21] Park, D.S., Chan, W., Zhang, Y., Chiu, C.C., Zoph, B., Cubuk, E.D. and Le, Q.V., 2019. Specaugment: A simple data augmentation method for automatic speech recognition. *arXiv preprint arXiv:1904.08779*.

[22] Stolcke, A., 2002. SRILM-an extensible language modeling toolkit. In *Seventh international conference on spoken language processing*.

Journal of Data Analysis
Vol.2, No.2, December 2019, p. 71-79

Parameter Estimation of the Temporal Point Process Model through the Bayesian Approach (Case study : Malaria disease data from Wahidin Hospital in Makassar City)

Darwis[1*], Sunusi N[2], Kresna A. J[3]

[1, 2, 3,]Mathematics Department, Faculty of Mathematics and Natural Sciences,
Hasanuddin University, Makassar
Email: awidarwis09@gmail.com*

Abstrak

Penelitian ini bertujuan mengestimasi parameter melalui pendekatan Bayesian dari model *temporal point process*. Paramater intensitas bersyarat model tersebut dipandang sebagai suatu *renewal process* yang selanjutnya digunakan melalui pendekatan *Squared Error Loss Function* (*SELF)*. Parameter intensitas bersyarat model *temporal point process* diestimasi menggunakan metode maximum likelihood estimation melalui persamaan *likelihood point process*. Selain itu, penelitian ini mengkaji metode estimasi maksimum likelihood dan metode Bayes untuk menganalis fungsi resiko dari hasil penaksir parameter intensitas bersyarat. Pada aplikasi estimasi parameter ini, studi kasus yang digunakan adalah menganalisa data orang yang terkena penyakit malaria yang datanya berasal dari Rumah Sakit Wahidin Kota Makassar. Studi kasus tersebut menghasilkan nilai $-1 \leq \theta < 0{,}4$ yang merupakan nilai resiko penaksir MLE yang lebih tinggi dibandingkan dengan menggunakan Metode Bayes sedangkan nilai $0.5 \leq \theta \leq 1$ merupakan hasil nilai resiko dari penaksir MLE yang lebih kecil dibandingkan dengan menggunakan Metode Bayes.

InformasiArtikel

Sejarah Artikel:
Diajukan 20 Jan 2020
Diterima 1 Apr 2020

Kata Kunci :
Temporal Point Process,
Penduga Bayes,
Sebaran Eksponensial

Abstract

This study aims to estimate parameters through the Bayesian approach from the temporal point process model. The conditional intensity parameter of the model is seen as an update process which is then used through the Squared Error Loss Function (SELF) approach. Generally the conditional intensity parameter for the temporal point process is estimated using the maximum likelihood estimation method that utilizes the likelihood point process equation but only limits the sample information extracted from a likelihood function without regard to prior information the sample. Another thing examined in this study is determining the Risk Function of the results of the conditional intensity parameter estimator using the maximum likelihood estimation (MLE) and the Bayes Method. As a parametric estimation application, an analysis of Malaria disease data was collected, the data obtained from Wahidin Hospital in

Keywords:
Temporal Point Process,
Bayesian Estimation,
Exponential Distribution

Makassar City. Based on the case study, the value is $-1 \leq \theta < 0.4$, MLE risk scale estimator is higher than risk scale made by estimator using Bayes while $0.5 \leq \theta \leq 1$, MLE risk scale estimator is smaller than estimator using Bayes method.

1. Introduction

Some research shows that climate change causes an increase in malaria incidence. Malaria is an infectious disease that can be transmitted by mosquitoes called Anopheles. This disease is most common in tropical and subtropical regions where the Plasmodium parasite can thrive as well as the Anopheles mosquito vector. Based on data in the world, malaria kills one child every 30 seconds. Around 300-500 million people are infected, and around one million people die from this disease every year [1]. Based on these cases, it is difficult for the community to predict when this malaria will befall them. Because of the difficulty in predicting the emergence of malaria, therefore, the authors try to solve the cases that occur.

In handling the case, the authors use the temporal point process conditional intensity parameter, which will be estimated using the Bayesian approach as done [2]. However, in this study, the intensity of the temporal point process is seen as a renewal process.

Therefore, this research will be written in a thesis entitled "Study temporal point process as a process of renewal (Renewal Process), Case Study Analyzing the time of emergence of malaria from data sourced from Wahidin Hospital in Makassar in 2011-2013".

2. Literature review

2.1 Temporal Point Process

A temporal point process is a model of point processes in one dimension (time) that is useful for sequences of random times if acase occurs. For example, the time when an earthquake occurs can be modeled as a temporal point process [3].

2.2 Renewal Process

The process of calculating the time between cases that are free and identical in a particular distribution is called the renewal process. The Renewal Process (N (t)) shows the number of cases during the interval [0, t]. Generally, X_n case is interpreted as the time between n (1-n) and to-n cases, which are commonly referred to as intercase time and S_n as the time until n-cases occur.

$$S_n = \sum_{i=1}^{n} X_i, \quad n \geq 1 \quad and \quad S_0 = 0 \tag{1}$$

2.3 Conditional Intensity Function

Conditional intensity function$\lambda(t|\mathcal{H}_t)$ is defined as P (one case in$(t, t+\Delta)|\mathcal{H}_t) = \lambda(t|\mathcal{H}_t)\Delta + o(\Delta)$, or

$$P(N(t, t+\Delta) = 1|\mathcal{H}_t) = \lambda(t|\mathcal{H}_t)\Delta + o(\Delta) \tag{2}$$

2.4 Risk Functions

Definition:
It is said that the L function is a loss function. The expected (average) value of the loss function is called the risk function. If $g(y;\theta), \theta \in \Omega$is the opportunity density function of y. Risk functions $R(\theta, \delta)$is as follows:

$$R(\theta, \delta) = E\{\mathcal{L}[\theta, \delta(y)]\} = \int_0^{\infty} \mathcal{L}[\theta, \delta(y)]g(y;\theta)dy \tag{3}$$

3. Reseach Methodology

In this research, the study of malaria time was taken. The data that will be used for this case study is malaria disease data sourced from Wahidin Sudirohusodo Hospital Makassar in the period 2011-2013. The variables used in this study are: the response variable in the study is malaria data that originated from Wahidin Sudirohusodo Hospital Makassar in the period 2011-2013.

1. Response variable :
   The response variable in the study was malaria disease data sourced from the Wahidin Sudirohusodo Hospital in Makassar.
2. Predictor variable :
   Time of Occurrence (Date, Month, and Year) and Time between events.

The implementation of this research begins with a theoretical review which is then applied to a real phenomenon, namely the emergence of a disease in a particular area. The steps taken are as follows:

1. State the conditional intensity of the temporal point process model as a renewal process.
2. Formulate the likelihood temporal point process function which is seen as a renewal process.
3. Determines the cumulative distribution function for the time distribution between events.
4. Define the prior sample distribution on the gamma distribution.
5. Formulate a posterior distribution.
6. Estimating conditional intensity parameters using the Bayes method.
7. Calculates the conditional intensity for the time between events having an exponential distribution.
8. The results of the conditional intensity parameter estimation obtained through the Bayesian approach are compared with the results of the parameter estimation using the MLE approach through the risk function

4. Results and Discussion

4.1 Renewal Process of Conditional Intensity Functions

As a renewal process, Conditional Intensity Function, the relationship between time between cases will be demonstrated in determining the conditional intensity function.

$$P(\text{nothing applied on } (t_{i-1}, t_i]) = exp\left(-\int_{t_{i-1}}^{t_i} \lambda(s|\mathcal{H}_s)ds\right) \quad (4)$$

4.2 Conditional Intensity of Time Between Cases with an Exponential Distribution

As is known the opportunity density function for exponential random variables, stated as

$$f(\tau_i) = \theta e^{-\theta(\tau_i)} \qquad , \tau_i \geq 0 dan\ 0 < \theta < \infty$$

and the cumulative density function is,

$$F(\tau_i) = 1 - e^{-\theta(\tau_i)} \qquad , \tau_i \geq 0$$

then the Conditional Intensity function:

$$\lambda\left(t_i|\mathcal{H}_{t_i}\right) = \frac{f(\tau_i)}{1 - F(\tau_i)},$$

$$= \frac{\theta e^{-\theta(\tau_i)}}{1 - (1 - e^{-\theta(\tau_i)})},$$

$$= \theta \, . \tag{5}$$

4.3 Bayesian Estimation of Exponential Distribution

Bayesian estimates for θ with the SELF approach are:

$$\hat{\theta}_s = \frac{n + \alpha}{\sum_{i=1}^{n} x_i + \frac{1}{\beta}} \tag{6}$$

where n =many cases, $\sum_{i=1}^{n} x_i$ = Amount of time inter-cases

$$\alpha = 1 \, , \; \beta = \frac{1}{\bar{x}}$$

4.4 Temporal Point Process Application

The plot data is obtained at Figure 1.

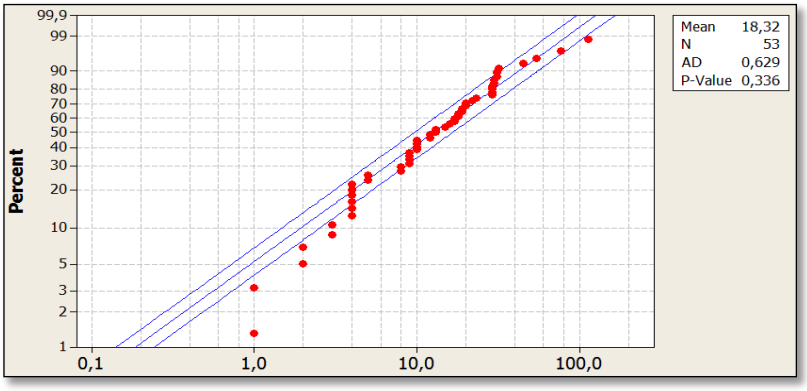


Figure 1 Probability The time plot inter-cases is exponentially distributed

Figure 1 shows the Probability of the test plot of the suitability of the exponential distribution with the time distribution between cases. The p-value and Anderson Darling (AD) can be seen in Table 1 as follows.

Table 1 Probability Value The time plot inter-cases has an Exponential distribution

Mean	N	AD	p-value
18.3	53	0.63	0.34

Table 1 It can be seen that the p-value from the Anderson Darling test above is higher than 0.34, which means it is greater than the alpha level of 0.05. So it is known that the data follows an exponential distribution.

4.5 Estimated Conditional Intensity Parameters

Table 2 Estimated Results of Conditional Intensity parameters for the time inter-cases with an Exponential distribution

Interval	Number of cases	Conditional Intensity
(0,1]	2	0.230531814
(1,2]	2	0.057632149
(2,3]	2	0.037459239
(3,4]	2	0.029378792
⋮	⋮	⋮

Interval	Number of cases	Conditional Intensity
(25,26]	3	0.004241373
(26,27]	2	0.003039283
(27,28]	1	0.001830873

4.6 MLE method

The parameter estimation results using the MLE method are obtained as follows: [4]

$$\lambda(t_i|\mathcal{H}_{t_i}) = \frac{n}{T} \quad , \text{ where T} = \sum x_i$$

Furthermore, the risk level of the parameter estimator using the MLE method will be determined as follows:

For example: $\sum x_i = y$

Noted that $L(\theta, \delta) = (\delta - \theta)^2$ as δ is an estimator for θ, then:

$$[(\delta - \theta)^2] = E[(\theta - \left(\frac{n}{y}\right)^2] \tag{7}$$

Based on equations (7),i.e.:

$$R(\theta, \delta) = E\{\mathcal{L}[\theta, \delta(y)]\} = \int_0^\infty \mathcal{L}[\theta, \delta(y)] g(y; \theta) dy$$

Where:

$$\mathcal{L}[\theta, \delta(y)] = \left[\theta - \left(\frac{n}{y}\right)\right]^2 .$$

Because of the sample $x_1, x_2 \ldots x_n$ having an exponential distribution then the corresponding peer prior distribution is the Gamma distribution so that:

$$R(\theta, \delta) = E\left[\theta - \left(\frac{n}{y}\right)\right]^2 = \int_0^\infty \left[\theta - \left(\frac{n}{y}\right)\right]^2 . \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} dy$$

$$= \int_0^\infty \left[\theta - \left(\frac{n}{y}\right)\right]^2 . \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} dy$$

$$= \int_0^\infty \left[\theta^2 - 2n\theta \frac{1}{y} + n^2 \frac{1}{y^2}\right] . \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} dy$$

$$= \int_0^\infty \theta^2 . \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} dy - \int_0^\infty 2n\theta \frac{1}{y} . \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} dy + \int_0^\infty n^2 \frac{1}{y^2} . \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} dy$$

$$= \theta^2 \int_0^\infty \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} dy - 2n\theta \int_0^\infty \frac{1}{y} . \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} dy + n^2 \int_0^\infty \frac{1}{y^2} \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} dy$$

$$= \theta^2 . 1 - \frac{2n\theta^{n+1}}{(n-1)!} \int_0^\infty y^{(n-1)-1} e^{-\theta y} dy + \frac{n^2 \theta^n}{(n-1)!} \int_0^\infty y^{(n-1)-2} e^{-\theta y} dy$$

$$= \theta^2 - \frac{2n\theta^{n+1}}{(n-1)!} . \int_0^\infty y^{(n-1)-1} e^{-\theta y} dy + \frac{n^2 \theta^n}{(n-1)!} . \int_0^\infty y^{(n-1)-2} e^{-\theta y} dy, \tag{8}$$

where :

$$\int_0^\infty y^{(n-1)-1} e^{-\theta y} dy = \int_0^\infty y^{(n-2)} e^{-\theta y} dy \tag{9}$$

and

$$\int_0^\infty y^{(n-1)-2} e^{-\theta y} dy \ = \int_0^\infty y^{(n-3)} e^{-\theta y} dy \, . \tag{10}$$

Equation (10) could be solved as follows :

Let: $\theta y = u$ thus $du = \theta dy$.

Journal of Data Analysis
Vol.2, No.2, December 2019, p. 71-79

Therefore$y = \frac{u}{\theta}$, thus $y^{n-2} = (\frac{u}{\theta})^{n-2}$,

then obtained $dy = \frac{du}{\theta}$.

Equation (10) can be written as:

$$\int_0^\infty y^{(n-2)}\, e^{-\theta y} dy = \int_0^\infty \left(\frac{u}{\theta}\right)^{n-2} e^{-u}.\frac{du}{\theta}$$

The results are obtained as follows:

$$\int_0^\infty y^{(n-2)}\, e^{-\theta y} dy \;= \int_0^\infty \frac{u^{n-2}}{\theta^{n-2}} e^{-u} \frac{1}{\theta} du$$

$$= \frac{1}{\theta}.\frac{1}{\theta^{n-2}} \int_0^\infty u^{n-2}\, e^{-u} du,$$

$$= \frac{1}{\theta}.\frac{1}{\frac{\theta^n}{\theta^2}} \int_0^\infty u^{n-2}\, e^{-u} du,$$

$$= \frac{1}{\theta}.\frac{\theta^2}{\theta^n} \int_0^\infty u^{n-2}\, e^{-u} du,$$

$$= \frac{\theta}{\theta^n} \int_0^\infty u^{n-2}\, e^{-u} du,$$

$$= \frac{\theta}{\theta^n}\,.(n-2)!$$

$$= \frac{(n-2)!}{\theta^{n-1}}. \qquad (11)$$

The same thing is done for Equation (11) so that it is obtained as follows:

$$\int_0^\infty \frac{1}{y^2} y^{(n-3)} e^{-\theta y} dy = \frac{(n-3)!}{\theta^{n-2}}\,. \qquad (12)$$

Based on Equation (11) and (12) obtained that:

$$R(\theta, \delta) = E\left[\theta - (\frac{n}{y})\right]^2 = \theta^2 - \frac{2n\theta^{n+1}}{(n-1)!} \cdot \int_0^\infty y^{(n-1)-1}\, e^{-\theta y} dy + \frac{n^2\theta^n}{(n-1)!} \cdot \int_0^\infty y^{(n-1)-2} e^{-\theta y} dy$$

$$= \theta^2 - \frac{2n\theta^{n+1}}{(n-1)!} \cdot \frac{(n-2)!}{\theta^{n-1}} + \frac{n^2\theta^n}{(n-1)!} \cdot \frac{(n-3)!}{\theta^{n-2}} = \theta^2 - \frac{2n\theta^2}{(n-1)} + \frac{n^2}{(n-1)(n-2)}\theta^2,$$

$$= \frac{(n-1)(n-2)\theta^2 - (n-2)2n\theta^2 + n^2\theta^2}{(n-1)(n-2)},$$

$$= \frac{(n^2-3n+2)\theta^2 - (2n^2-4n)\theta^2 + n^2\theta^2}{(n-1)(n-2)},$$

$$= \frac{n^2\theta^2 - 3n\theta^2 + 2\theta^2 - 2n^2\theta^2 + 4n\theta^2 + n^2\theta^2}{(n-1)(n-2)},$$

$$= \left(\frac{(n+2)}{(n-1)(n-2)}\right)\theta^2\,. \qquad (13)$$

4.7 Bayes methods

Noticed that :

$$\hat{\theta}_s = \frac{n+\alpha}{\sum_{i=1}^{n} x_i + \frac{1}{\beta}}$$

The Equation can be written as follows:

$$\frac{n+\alpha}{\sum_{i=1}^{n} x_i + \frac{1}{\beta}} = \frac{n}{y+\frac{1}{\beta}} + \frac{\alpha}{y+\frac{1}{\beta}}$$
$$= \frac{n}{y}\left(\frac{y}{y+\frac{1}{\beta}}\right) + \alpha\beta\left(\frac{1}{\left(y+\frac{1}{\beta}\right)\beta}\right).$$

Let

$$A = \frac{y}{y+\frac{1}{\beta}} \text{ and } B = \frac{1}{\left(y+\frac{1}{\beta}\right)\beta} = \frac{1}{y\beta+1}$$
$$A = \frac{y\beta}{y\beta+1} \ , \ B = 1 - A$$

Thus

$$\frac{n+\alpha}{\sum_{i=1}^{n} x_i + \frac{1}{\beta}} = \frac{n}{y+\frac{1}{\beta}} + \frac{\alpha}{y+\frac{1}{\beta}}$$
$$= \frac{n}{y}\left(\frac{y}{y+\frac{1}{\beta}}\right) + \alpha\beta\left(\frac{1}{\left(y+\frac{1}{\beta}\right)\beta}\right)$$
$$= \frac{n}{y}.A + \alpha\beta.B.$$

Since $E(\theta) = \hat{\theta}$ and $\hat{\theta} = \frac{n+\alpha}{\sum_{i=1}^{n} x_i + \frac{1}{\beta}}$,

then

$$= \frac{n+\alpha}{\sum_{i=1}^{n} x_i + \frac{1}{\beta}} = \frac{n}{y}.A + \alpha\beta.B \ .$$

Furthermore

$$R(\theta,\delta) = E\left[\left[\theta - \left(\frac{n+\alpha}{y+\frac{1}{\beta}}\right)\right]^2\right] = E\left[\left[\theta - \frac{n}{y}A + (B)\alpha\beta\right]^2\right],$$

or

$$E\left[\left[\theta - \left(A\frac{n}{y} + (1-A)\alpha\beta\right)\right]^2\right]$$
$$= E\left[\theta^2 - 2\theta\left(A\frac{n}{y}\right) - 2\theta(1-A)\alpha\beta - 2A(1-A)\frac{n}{y}\alpha\beta + ((1-A)\alpha\beta)^2 + A^2\left(\frac{n}{y}\right)^2\right],$$
$$= E(\theta^2) - E\left(2A\theta\frac{n}{y}\right) - E(2\theta(1-A)\alpha\beta) - E\left(2A(1-A)\frac{n}{y}\alpha\beta\right) + E\left(((1-A)\alpha\beta)^2\right) + E\left(A^2\left(\frac{n}{y}\right)^2\right)$$
$$= \theta^2 - 2A\theta n\, E\left(\frac{1}{y}\right) - 2\theta(1-A)\alpha\beta - 2A(1-A)\alpha\beta n\, E\left(\frac{1}{y}\right) + ((1-A)\alpha\beta)^2 + A^2n^2E\left(\frac{1}{y^2}\right),$$
$$= \theta^2 - 2A\theta\frac{n}{n-1}\theta - 2\theta(1-A)\alpha\beta - 2A(1-A)\alpha\beta\ \frac{n}{n-1}\theta + (1-A)\alpha\beta)^2 + A^2n^2\frac{\theta^2}{(n-1)(n-2)},$$
$$= \left(1 - 2A\frac{n}{n-1} + \frac{A^2n^2}{(n-1)(n-2)}\right)\theta^2 - \left(2(1-A)\alpha\beta + 2A(1-A)\alpha\beta\frac{n}{n-1}\right)\theta + (1-A)\alpha\beta,$$

$$= \left(\frac{n^2(1-2A+A^2)-(3-4A)n+2}{(n-1)(n-2)}\right)\theta^2 - \left(\frac{(2A\alpha\beta-2A^2\alpha\beta)n+2(2-2A)\alpha\beta}{(n-1)}\right)\theta + (1-A)(\alpha\beta)^2 \quad (14)$$

4.8 The differences in the risk functions of the MLE and Bayes methods

Based on the case study, the risk function is obtained as follows :
For Bayes method,

$$R(\theta,\hat{\theta}) = 0.02065\theta^2 + 0.00217\theta - 0.00102$$

For,

$$\text{MLE}R(\theta,\hat{\theta}) = 0.02073\theta^2$$

Table 3 The score of risk estimator scale for the of MLE method and Bayes for the score $-1 \leq \theta \leq 1$

Scoreθ	Risk Scale		Scoreθ	Risk Scale	
	MLE	BAYES		MLE	BAYES
-1	0.020739	0.01746	0	0	-0.00102
-0.9	0.016799	0.01375	0.1	0.000207	-0.00060
-0.8	0.013273	0.01046	0.2	0.00083	0.00024
-0.7	0.010162	0.00758	0.3	0.001867	0.00149
-0.6	0.007466	0.00511	0.4	0.003318	0.00315
-0.5	0.005185	0.00306	0.5	0.005185	0.00523
-0.4	0.003318	0.00142	0.6	0.007466	0.00772
-0.3	0.001867	0.00019	0.7	0.010162	0.01062
-0.2	0.00083	-0.00063	0.8	0.013273	0.01393
-0,1	0.000207	-0.00103	0.9	0.016799	0.01766
0	0	-0.00102	1	0.020739	0.02180

Table 3 indicatesthat: $-1 \leq \theta < 0.4$, MLE risk scale estimator is higher than risk scale made by estimator using Bayes while $0.5 \leq \theta \leq 1$, MLE risk scale estimator is smaller than estimator using Bayes method.

It can be seen in the graph below :

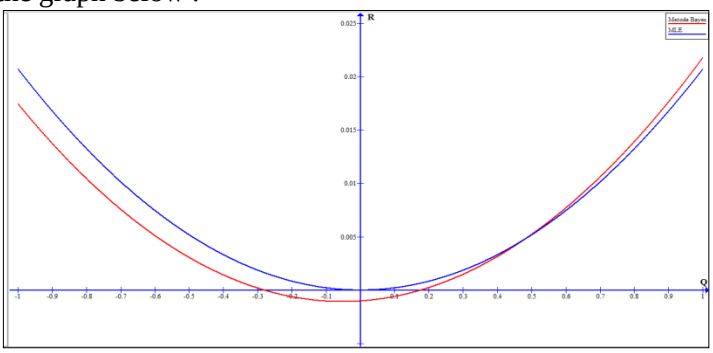


Figure 2 Graph of the relationship between parameter score $\theta$with risk scale using MLE and Bayes methods

Note:

$Q = \theta$and R =Risk Scale

Figure 2 shows that score $-1 \leq \theta \leq 1$, MLE estimator (blue line) presents a higher risk of error than the risk made by estimator using the Bayes method (red line).

5. Conclusion

Based on the description from the previous chapters, it can be concluded as follows: The estimated conditional intensity parameters of the temporal point process model through the Bayesian approach are carried out as follows:

First, stating the conditional intensity time between cases, then constructing the likelihood temporal point process:

$$L = \left(\prod_{i=1}^{N} f\left(t_i|\mathcal{H}_{t_i}\right)\right) P\left(N_{(t_N,T]} = 0\right),$$
$$= \left(\prod_{i=1}^{N} \lambda\left(t_i|\mathcal{H}_{t_i}\right)\right) exp\left\{-\int_0^T \lambda(s|\mathcal{H}_s)dt\right\},$$

Furthermore, the estimation of conditional intensity parameters using the Bayesian approach is done through the SELF approach. Estimation results obtained as follows:

$$\lambda\left(t_i|\mathcal{H}_{t_i}\right) = \frac{n+\alpha}{\sum_{i=1}^{n} x_i + \frac{1}{\beta}}.$$

1. Risk function of each estimator:

Bayes estimator :

$$R\left(\theta,\hat{\theta}\right) = \left(\frac{n^2(1-2A+A^2)-(3-4A)n+2}{(n-1)(n-2)}\right)\theta^2 \quad -\left(\frac{(2A\alpha\beta-2A^2\alpha\beta)n+2(2-2A)\alpha\beta}{(n-1)}\right)\theta + (1-A)(\alpha\beta)^2 \ .$$

MLE estimator :

$$R\left(\theta,\hat{\theta}\right) = \left(\frac{(n+2)}{(n-1)(n-2)}\right)\theta^2 .$$

2. From the data recorded at Wahidin Sudirohusodo Hospital illustration is obtained:

MLE estimator score presents a higher error than the risk that made by estimator using the Bayes method.

6. Recommendation

In this study, the study of the emergence of malaria for the foreseeable future can not be predicted because research is still limited to determining conditional intensity. It is recommended in subsequent studies to make parametric methods for the conditional intensity that has been obtained.

References

[1] Sumampouw O.J, dkk (2019), "the trend of dengue hemorrhagic fever in the city of Manado in 2018-2019.*KESMAS.*8(6).

[2] Ogata.(1999),*"Seismicity analysis through point process modeling :A review*.Pure Appl. Geophys. 155,471-507.

[3] Daley, D.J. and Vere-Jones, D. (2003) "*Introduction to the Theory of Point Processes*". Vol 1: Elementary Theory and Methods(2nd edition). Springer, New York.

[4] Sunusi, N,dkk , (2010), " Study of Eartquake Forecast through Hazard Rate Analysis, *Int.J.App*. Math and Stat, vol 1(J.10) : p. 96-103

Journal of Data Analysis
Vol.2, No.2, December 2019, p. 80-87

Analisis Kepuasan Pengguna Aplikasi RWikiStat 3.0

Hizir Sofyan[1], Rasudin[2], Miftahuddin[3], Kurnia Saputra[4], Marzuki[5*], Muhammad Iqbal[6], Doddy Maulana[7]

[1,3,5,6]Statistika, FMIPA, Universitas Syiah Kuala, Banda Aceh
[2,4,7]Informatika, FMIPA, Universitas Syiah Kuala, Banda Aceh
Email: hizir@unsyiah.ac.id; rasudin@unsyiah.ac.id; miftahuddin@unsyiah.ac.id; kurnia.saputra@unsyiah.ac.id; marzuki@unsyiah.ac.id*; m.iqbal@mhs.unsyiah.ac.id; doddym0@gmail.com

Abstrak

RWikiStat 3.0 adalah aplikasi android untuk pebelajaran statistika berbasis RWeb dan Teknologi Wiki. Aplikasi ini merupakan pengembangan dari RWikiStat 2.0. Kepuasan pengguna aplikasi RWikiStat 3.0. dianalisis dalam tulisan ini. Performa yang dianalisis adalah tampilan aplikasi, tingkat responsif, dan kemanfaatan aplikasi. Data penelitian diperoleh dengan metode survei. Survei dilakukan setelah pelatihan penggunaan aplikasi ini. Pelatihan tersebut dilakukan pada tiga perguruan tinggi di Banda Aceh, yaitu Universitas Serambi Mekkah (USM), Universitas Syiah Kuala (Unsyiah), dan Sekolah Tinggi Keguruan dan Ilmu Pendidikan Bina Bangsa Getsempena (STKIP BBG). Sampel diambil dengan menggunakan metode cluster random sampling dan Unsyiah terambil sebagai klaster penelitian. Jumlah sampel dari Unsyiah adalah sebanyak 37 responden. Responden diberikan angket yang mengandung 9 pertanyaan terkait dengan pendeskripsian secara umum tentang kepuasan responden sebagai pengguna terhadap aplikasi RWikiStat 3.0. Hasil analisis menghasilkan bahwa secara umum responden telah puas akan aplikasi RWikiStat 3.0. Kepuasan terhadap tampilan aplikasi, tingkat responsif, dan kebergunaan aplikasi cukup tinggi. Kemudian, responden memiliki keinginan yang besar untuk merekomendasikan aplikasi ke teman, kolega, atau lainnya.

Informasi Artikel

Sejarah Artikel:
Diajukan 9 Maret 2020
Diterima 8 April 2020

Kata Kunci :
RWikiStat
Android
Pembelajaran Statistika
Kepuasan
Pengguna Aplikasi
RWeb

Abstract

RWikiStat 3.0 is an android application for RWeb and Wiki technological -based statistics learning. This application is the development of RWikiStat 2.0. User satisfaction of RWikiStat 3.0 application was analysed in this paper. Performance was analysed based on the application interface, responsiveness, and features of the application. The research data were obtained by survey method conducted after the training to use this application. The training was conducted at three universities in Banda Aceh, namely Serambi Mekkah University (USM), Universitas Syiah Kuala (Unsyiah), and the Higher School of Teacher Training and Education of Bina Bangsa Getsempena (STKIP BBG). Samples were taken using cluster random sampling method and Unsyiah was fetched as research clusters.

Keywords:
RWikiStat
Android
Statistical Learning
Satisfaction
Application Users
RWeb

The number of samples of Unsyiah were 37 respondents. Respondents were given a questionnaire containing nine questions related to the general description of respondent satisfaction as users for using the application of RWikiStat 3.0. The results of the analysis showed that the overall respondents were satisfied using the application of RWikiStat 3.0. There was higher satisfaction in the application interface, the level of responsiveness and usability of applications. Then, the respondents had a great intention to recommend the application to friends, colleagues, or others.

1. Pendahuluan

R merupakan salah satu perangkat lunak dalam pemrograman, terutama dalam analisis statistik. Perangkat lunak ini semakin populer karena dapat melakukan analisis statistik dengan berbagai metode statistik, seperti statistik deskriptif, statistik inferensia, dan big data [1]. R merupakan perangkat lunak berbasis *open source* yang dapat diunduh dan dipasangkan pada komputer [2]. Saat ini R telah dikembangkan untuk berbagai sistem operasi seperti Linux, Windows and Macintosh.

Penggunaan perangkat lunak R saat ini banyak pada komputer. Sedangkan untuk perangkat telepon selular masih belum banyak. Penggunaan perangkat tersebut sangat berkembang pesat untuk saat ini sehingga diperlukan pengembangan penggunaan perangkat lunak R ini pada perangkat telepon selular.

Telepon selular sangat praktis digunakan karena mudah dibawa-bawa. Banyak aplikasi yang dapat digunakan dalam telepon selular, di antara yang sedang berkembang pesat saat ini adalah aplikasi android. Sistem operasi android bersifat terbuka yang dikembangkan oleh Google. Android merupakan sistem operasi berbasis Java.

RWikiStat 3.0 merupakan aplikasi R berbasis android yang dikembangkan oleh peneliti dari Universitas Syiah Kuala. RWikiStat 3.0 merupakan pengembangan dari versi RWikiStat 1.0 dan RWikiStat 2.0. Aplikasi ini diharapkan memudahkan pengerjaan metode statistik dimanapun dan kapanpun saat diperlukan.

Penelitian ini bertujuan untuk melihat kepuasan pengguna aplikasi RWikiStat 3.0. Kepuasan yang dianalisis adalah tampilan aplikasi, tingkat responsif, dan kemanfaatan aplikasi. Tulisan ini diharapkan dapat digunakan untuk peningkatan kinerja aplikasi RWikiStat sehingga semakin mempermudahkan bagi pengguna dalam memanfaatkan aplikasi ini di bidang statistika.

2. Tinjauan Kepustakaan

Rweb adalah web based interface untuk R (*a statistical analysis package*) dengan men-submit kode, menjalankan R kode (dalam *batch mode*), dan output-nya dalam bentuk *printed* dan *graphical*. Perintah-perintah Rweb adalah identik dengan perintah-perintah di R. Rweb dikembangkan oleh Jeff Banfield dari *Department of Mathematical Sciences, University of Montana, USA.* Paket-paket pendukung yang diperlukan untuk menjalankan Rweb server di antaranya R beserta dengan dokumentasinya, source code, R libraries, dan beberapa paket-paket pendukung dalam R, Perl, netpbm. NetPBM library dibutuhkan untuk mengkonversikan berbagai macam format grafik, LWP Perl module, CGI Perl module, dan ghostscript.

Android merupakan sistem operasi mutakhir yang mendukung jumlah aplikasi yang sangat banyak pada ponsel pintar. Android menyediakan platform dan sistem operasi untuk perangkat

mobile yang bernuansa Linux dan dikembangkan oleh Google dan Open Handset Alliance. Bahasa pemrograman yang digunakan oleh Android adalah bahasa pemrograman Java. Android merolusikan lingkungan perangkat *mobile*. Untuk pertama kalinya Android merupakan *platform* terbuka yang memisahkan antara perangkat kerasnya dari perangkat lunak yang menjalankannya. Hal ini memungkinkan untuk menambah jumlah perangkat yang sangat besar untuk menjalankan aplikasi yang sama dan menciptakan ekosistem yang kaya bagi pengembang maupun pengguna.

Sistem operasi Android memiliki arsitektur yang berupa lapisan. Lapisan-lapisan ini memiliki karakteristik dan tujuan tersendiri. Lapisan pada sistem operasi Android ada 5 yaitu lapisan aplikasi, lapisan *framework* aplikasi, lapisan *libraries*, lapisan *runtime* Android, dan lapisan kernel Linux. Lapisan aplikasi terdiri atas himpunan aplikasi inti seperti klien email, program SMS, kalender, peta, *browser*, kontak, dan lain-lain. Semua aplikasi dibangun dengan menggunakan bahasa pemrograman Java. Tiap aplikasi bertujuan untuk menjalankan tugas tersendiri. Menurut Andy Rubin, pendiri Android Inc., Android merupakan platform yang disediakan untuk perangkat *mobile* yang sederhana namun menyeluruh dan bersifat terbuka sehingga inovasi yang ada dapat tersalurkan dengan cepat tanpa harus memikirkan lisensi.

RWikiStat adalah aplikasi pembelajaran statistik berbasis Rweb. Aplikasi ini dibangun dengan menggunakan Rweb (pembelajaran berbasis web untuk perangkat lunak statistik R) dan Teknologi Wiki (MediaWiki). Rweb dianggap sebagai penghubung antara konten dan pengguna dalam proses pembelajaran dan MediaWiki adalah untuk menjamin keberlanjutan program.

Kajian mengenai pembelajaran Statistika berbasis RWikiStat telah dirintis sejak tahun 2004 [3]. Kemudian kajian mengenai rancang-bangun pembelajaran interaktif melalui R dilakukan pada tahun 2008. Selanjutnya kajian tentang RWikiStat 1.0 dengan bantuan CD pada tahun 2010 [4] dan RWikiStat 2.0 pada tahun 2012 [5]. Aplikasi ini berkembang seiring dengan kemajuan teknologi. Setiap produk yang dihasilkan diperlukan evaluasi yang salah satunya dari sisi pengguna produk seperti aplikasi ini. Analisis kepuasan pengguna merupakan salah satu analisis statistika yang dapat digunakan dalam evaluasi tersebut. Salah satu kepuasan pengguna dapat diketahui melalui kuesioner. Kuesioner adalah salah satu cara pengumpulan data yang dilakukan dengan memberi sejumlah pertanyaan tertulis untuk dijawab oleh responden [6]. Pertanyaan-pertanyaan ini dimaksudkan untuk memperoleh informasi tentang dimensi-dimensi kualitas perlayanan.

Penelitian tentang kepuasan yang dianalisis dengan deskriptif sudah sering dilakukan oleh peneliti. Salah satunya adalah analisis deskriptif kepuasan karyawan lembaga keuangan perbankan. Penelitian yang dilakukan di Surabaya ini mengambil beberapa indikator yang dijadikan pertanyaan kepada karyawan [7]. Penelitian ini dilakukan tahun 2017.

3. Metode Penelitian

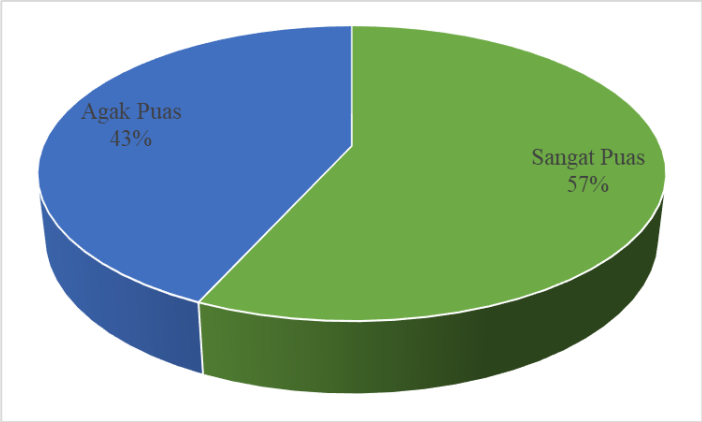
Data penelitian diperoleh dengan metode survei. Survei ini dilakukan untuk mengetahui kepuasan pengguna terhadap aplikasi RWikiStat 3.0. Survei dilakukan setelah pelatihan penggunaan aplikasi ini. Sebelumnya, pelatihan tersebut telah dilakukan pada tiga perguruan tinggi di Banda Aceh, yaitu Universitas Serambi Mekkah (USM), Universitas Syiah Kuala (Unsyiah), dan Sekolah Tinggi Keguruan dan Ilmu Pendidikan Bina Bangsa Getsempena (STKIP BBG). Sampel diambil dengan menggunakan metode *cluster random sampling* dengan ketiga perguruan tinggi sebagai klaster. Setelah dilakukan pengacakan, Unsyiah terambil sebagai klaster penelitian. Jumlah sampel dari Unsyiah adalah sebanyak 37 responden atau sekitar sepertiga dari populasi.

Analisis dalam penelitian ini menggunakan analisis deskriptif. Responden diberikan angket mengenai kepuasan terhadap aplikasi RWikiStat 3.0. Angket ini mengandung 9 pertanyaan. Pertanyaan yang diberikan terkait dengan pendeskripsian secara umum tentang kepuasan responden sebagai pengguna terhadap aplikasi RWikiStat 3.0. Aplikasi ini dapat diunduh pada laman web Jurusan Statistika Universitas Syiah Kuala, https://stat.unsyiah.ac.id.

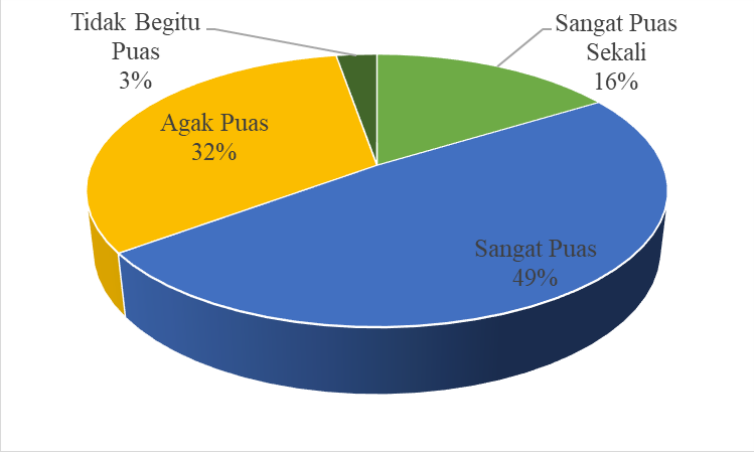
4. Analisis Pembahasan

Responden terdiri dari 43% berjenis kelamu laki-laki dan 57% perempuan. Responden merupakan mahasiswa dari 6 angkatan yaitu 24 orang angkatan 2019, 7 dari angkatan 2018, masing-masing 2 orang dari angkatan 2017 dan 2015, serta masing-masing 1 orang merupakan mahasiswa angkatan 2016 dan 2013.

Semua responden puas terhadap aplikasi RWikiStat 3.0, dengan rincian sebanyak 57% responden sangat puas dan 43% agak puas terdapat aplikasi RWikiStat 3.0. Perbandingan kepuasan terhadap aplikasi ini secara jelas dapat dilihat pada Gambar 1.

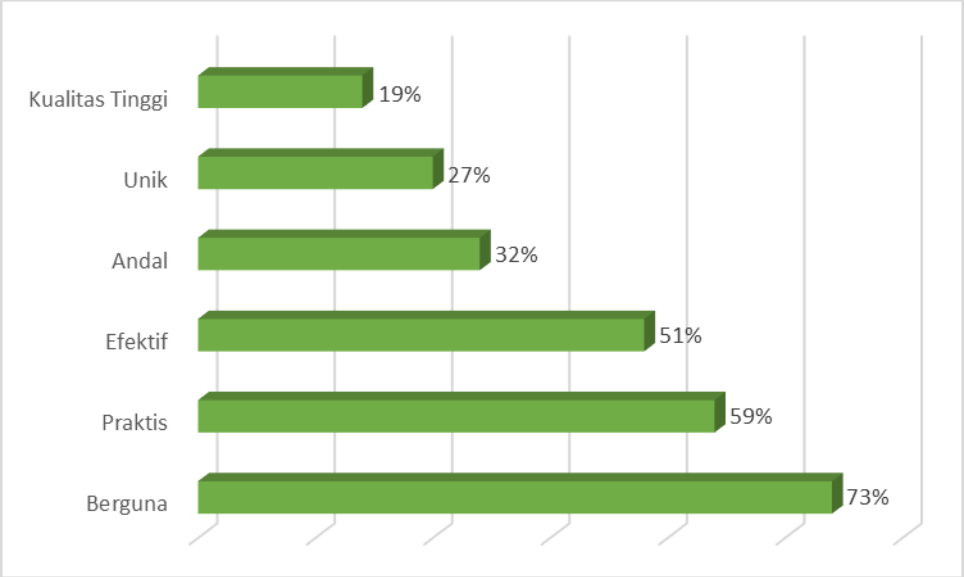


Gambar 1 Kepuasan Aplikasi RWikiStat 3.0 secara keseluruhan



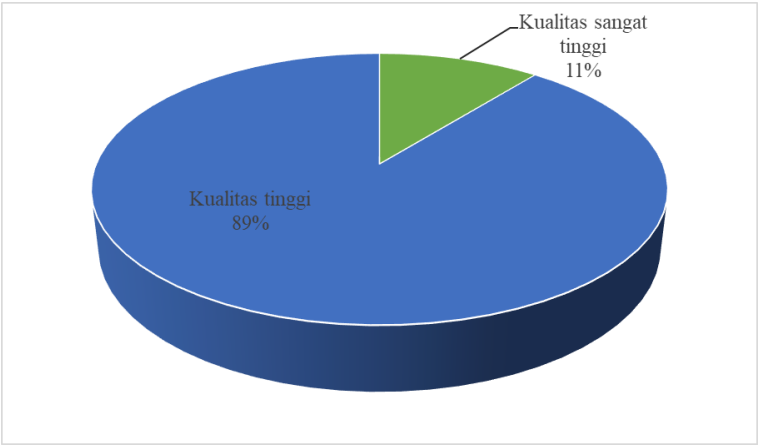
Gambar 2 Tampilan dan nuansa aplikasi RWikiStat 3.0

Tampilan dan nuansa aplikasi RWikiStat 3.0. dinilai oleh 97% pengguna dengan sangat puas. Sedangkan 3% lainnya tidak begitu puas dengan tampilan dan nuansa aplikasi ini. Gambar 2 menampilkan perbandingan pendapat responden terhadap tampilan aplikasi.



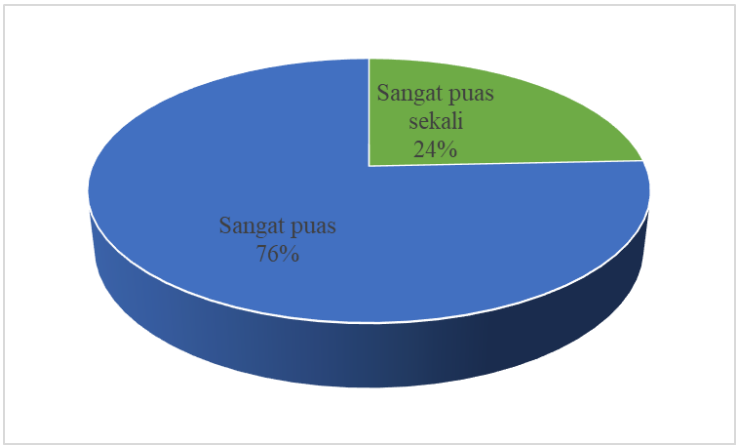
Gambar 3 Kata-kata berikut yang menggambarkan aplikasi RWikiStat 3.0.

Responden yang menyatakan bahwa aplikasi RWikiStat 3.0 ini adalah berguna adalah sebanyak 73%, praktis (59%) efektif (51%), andal (32%), unik (27%), dan memiliki kualitas tinggi (19%). Gambar 3 menyajikan persentase responden dalam menggambarkan aplikasi RwikiStat 3.0.



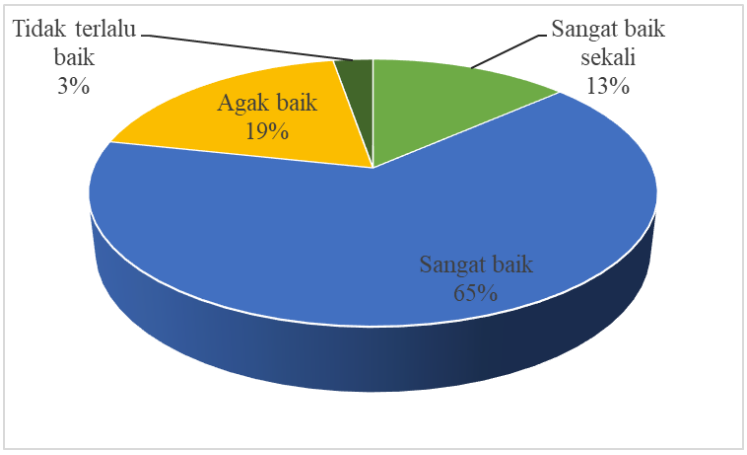
Gambar 4 Nilai kualitas aplikasi RWikiStat 3.0

Responden yang menyatakan bahwa aplikasi RWikiStat 3.0 berkualitas tinggi sebanyak 19%. Dari 19% responden tersebut, bahkan 11% di antaranya menilai aplikasi ini berkualitas sangat tinggi (Gambar 4).



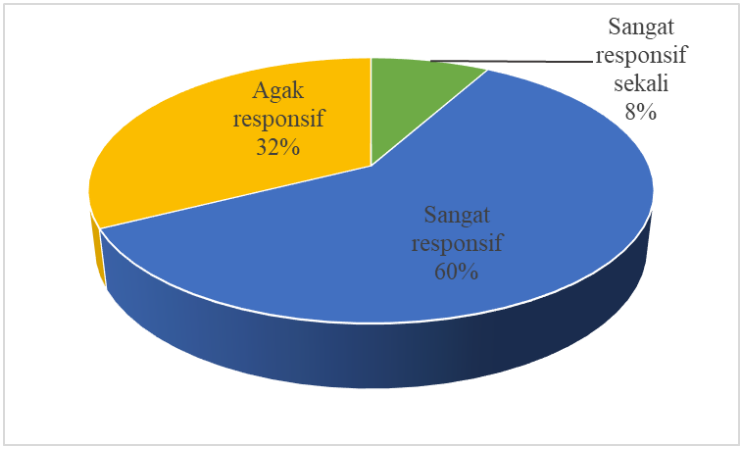
Gambar 5 Keandalan aplikasi RWikiStat 3.0

Aplikasi ini dinyatakan andal oleh 32% responden, dengan rincian sebanyak 76% responden puas dengan keandalan aplikasi ini dan sisanya yaitu 24% menyatakan sangat puas sekali.



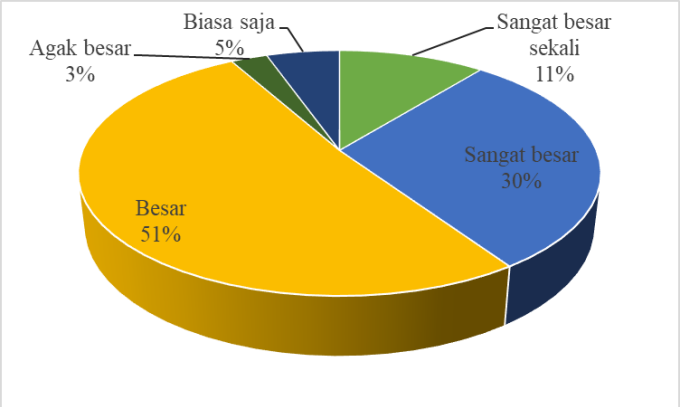
Gambar 6 Aplikasi RWikiStat 3.0 memenuhi kebutuhan pengguna

Sebanyak 97% responden mengatakan aplikasi RWikiStat 3.0 memenuhi kebutuhan pengguna, sedangkan 3% lainnya tidak begitu baik memenuhi kebutuhan.



Gambar 7 Aplikasi RWikiStat 3.0 terhadap pertanyaan atau masalah pengguna

Sebanyak 68% responden mengatakan aplikasi RWikiStat 3.0 sudah responsif, dan 32% lainnya mengatakan agak responsif.



Gambar 8 Merekomendasikan aplikasi RWikiStat 3.0 ke teman, kolega, atau lainnya

Sebanyak 92% responden, besar keinginan untuk merekomendasikan aplikasi RWikiStat 3.0 ke teman, kolega, atau lainnya. Sedangkan 3% lainnya agak besar, dan 5% sisanya biasa saja.

Responden juga mengharapkan agar tampilan dari RWikiStat 3.0 lebih ditingkatkan. Selain itu, diperlukannya perbaikan dalam command line, sehingga lebih mudah dalam menjalankan sintak-sintak R programming.

5. Kesimpulan dan Saran

Analisis terhadap kepuasan pengguna aplikasi RWikiStat menghasilkan bahwa secara umum responden telah puas akan aplikasi RWikiStat 3.0. Hal ini dapat dilihat dari tingginya kepuasan terhadap tampilan aplikasi, tingkat responsif, dan kebergunaan aplikasi. Selain itu, responden juga memiliki keinginan yang besar untuk merekomendasikan aplikasi ke teman, kolega, atau lainnya.

Terima Kasih

Terima kasih kami ucapkan kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Syiah Kuala yang telah mendukung kegiatan pengabdian dan publikasi ini dan kepada Jurusan Statistika Universitas Syiah Kuala yang telah mendukung kegiatan pelatihan.

Daftar Pustaka

[1] R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, 2009, URL http://www.R-project.org.

[2] http://www.math.montana.edu/Rweb/. Last access in January 2010.

[3] Sofyan, H., Achsani, N.A., 2004, MM*INDO : Interactive Statistics Learning In Indonesian Language, STATISTIKA: Forum Teori dan Aplikasi Statistika, 4, p. 1411-5891.

[4] Subianto, M., Sofyan, H. 2010. Interactive Statistics Learning with RWikiStat. Proceedings of the 2010 International Conference on Networking and Information Technology (ICNIT). Manila : IEEE.

[5] Sofyan, H., Muttaqin, E., Subianto, M. 2012. RWikiStat 2.0: a Web Based Statistical Learning System. Proceedings of the Symposium of Japanese Society of Computational Statistics 26. Tokyo : Japanese Society of Computational Statistics.

[6] Christsanto, J., Negoro, N.P., Rahmawati, Y. 2017. Analisis Deskriptif Kepuasan Karyawan Lembaga Keuangan Perbankan. Jurnal Sains dan Seni ITS 6(2). Surabaya : ITS.

[7] Sugiyono. 2006. Metode Penelitian Bisnis. Bandung : Alfabeta.

Journal of Data Analysis
Vol.2, No.2, December 2019, p. 88 - 95

Penerapan Persamaan Model Struktural dalam Mengidentifikasi Variabel yang dapat Mempengaruhi Status Gizi Remaja di Kabupaten Aceh Besar

Latifah Rahayu[1*], Samsul Anwar[2], Winny Dian Safitri[3] dan Radian Akrama[4]

[1,2,3,4] Department of Statistics, Syiah Kuala University, Banda Aceh, Indonesia
E-mail: latifah.rahayu@unsyiah.ac.id*

* = corresponding author

Abstrak

Structural Equation Model (SEM) biasanya digunakan pada pengujian rangkaian hubungan variabel yang relatif sulit diukur secara bersamaan. Rangkaian ini merupakan hubungan yang terbentuk dari satu atau lebih variabel bebas dengan atau lebih dari satu variabel tak bebas. Bidang kesehatan merupakan salah satu bidang penelitian yang banyak menerapkan metode SEM. Hal ini mengingat bahwa pada bidang kesehatan, banyak terdapat variabel laten atau variabel yang tidak dapat diukur secara langsung. Tujuan dilakukannya penelitian ini adalah untuk mengetahui variabel apa saja yang dapat mempengaruhi status gizi remaja di Kabupaten Aceh Besar. Variabel laten berupa status gizi remaja sedangkan variabel bebas berupa kebiasaan dan pola makan, status dan kondisi kesehatan, kondisi keluarga, dan pengetahuan remaja. Berdasarkan hasil model SEM yang dibentuk, beberapa faktor yang memengaruhi status gizi remaja di Kabupaten Aceh Besar adalah variabel status dan kondisi kesehatan, dan variabel kondisi keluarga.

Informasi Artikel

Sejarah Artikel:
Diajukan 24 April 2020
Diterima 3 Mei 2020

Kata Kunci:
Structural Equation Model
Analisis Jalur
Status Gizi Remaja

Abstract

Structural Equation Model (SEM) is apply to test a series of relationships that are relatively difficult to measure simultaneously. This relationship is a relationship that is formed from one or more independent variables with or more than one independent variable. The health sector is one of the research fields that widely applies the SEM method. This is considering that in the health sector, there are many latent variables or variables that cannot be measured directly. This research was conducted with the objective of finding out what factors influenced the nutritional status of adolescents in Aceh Besar District. The latent variable is in the form of adolescent nutritional status while the independent variable is in the form of habits and eating patterns, health status and the conditions, family conditions, and adolescent knowledge. Based on the results of the SEM method, the factors that influence the nutritional status of adolescents in Aceh Besar District are health status and condition, and family conditions.

Keyword:
Structural Equation Model
Path Analysis
Adolescent Nutritional Status

Journal of Data Analysis
Vol.2, No.2, December 2019, p. 88 - 95

1. Pendahuluan

Bentuk dari *Structural Equation Model* (SEM) merupakan gabungan dari beberapa metode statistik yang dapat menguji suatu ikatan jaringan yang dibentuk antara satu variabel atau beberapa variabel terikat dengan satu variabel atau beberapa variabel bebas, di mana pada setiap variabel terikat dan bebas berbentuk bangunan yang dibentuk dari beberapa parameter yang diukur secara bersamaan[1]. Di mana variabel-variabel tersebut dapat berupa variabel yang tidak dapat diukur (laten) atau yang dapat diukur (teramati). Metode SEM biasanya digunakan untuk menguji ikatan hubungan yang relatif sulit diukur secara bersamaan.

Bidang kesehatan merupakan salah satu bidang penelitian yang banyak menerapkan metode SEM. Dewasa ini, permasalahan status gizi remaja di Aceh adalah isu kesehatan yang membutuhkan perhatian khusus dari segala pihak terutama pemerintah. Hal ini ditunjukkan dari hasil penelitian mengenai gizi remaja yang telah dilakukan oleh [2] dengan tujuan mengetahui hubungan konsumsi makanan cepat saji dengan status gizi pada pelajar Sekolah Menengah Atas yang ada di Kabupaten Aceh Besar. Hasil penelitian tersebut menunjukkan bahwa permasalahan status gizi merupakan permasalahan kesehatan pada remaja di Kabupaten Aceh Besar. Persentase remaja dengan gizi kurang dan kegemukan di Kabupaten Aceh Besar masing-masing sebesar 24,3%, sedangkan remaja dengan masalah status gizi obesitas hanya sebesar 2,4%. Secara umum, remaja yang mengalami masalah status gizi remaja di Kabupaten Aceh Besar adalah sebanyak 51,2%.

Penelitian yang juga berkaitan dengan status gizi remaja telah dilakukan oleh [3]dengan tujuan mengidentifikasi status gizi remaja di Kota Banda Aceh yang bersebelahan langsung dengan Kabupaten Aceh Besar. Hasil penelitian tersebut menunjukkan bahwa sebanyak 0,33% remaja mengalami gizi buruk, 3,48% remaja mengalami gizi kurang, 65,17% remaja dalam kategori gizi baik dan 15,59% remaja berstatus obesitas. Dengan melihat beberapa penelitian terdahulu, maka perlu dilakukan identifikasi terhadap variabel-variabel yang memengaruhi status gizi remaja terutama di Kabupaten Aceh Besar menggunakan Metode *Structural Equations Model* (SEM).

2. Tinjauan Kepustakaan

2.1 Persamaan pada SEM

Secara umum, bentuk persamaan SEM terdiri dari persamaan model struktural dan persamaan model pengukuran. Menurut [4], model struktural merupakan model yang menggambarkan hubungan antar variabel laten yang dibentuk berdasarkan substansi teori. Berikut adalah persamaan untuk model struktural :

$$\eta_1 = \gamma_{11}\xi_1 + \zeta_1 \quad (1)$$

$$\eta_2 = \gamma_{22}\xi_1 + \zeta_1 \quad (2)$$

$$\eta_3 = \beta_{31}\eta_1 + \beta_{32}\eta_2 + \zeta_3 \quad (3)$$

Di mana

ξ (*ksi*)	:	Konstruk laten eksogen
η (*eta*)	:	Konsruk laten endogen
β (*beta*)	:	parameter yang dapat menggambarkan hubungan secara langsung antara variabel endogen dengan variabel endogen lainnya
ζ (z*eta*)	:	kesalahan struktural (*structural error*) yang terdapat pada konstruk endogen.

Model pengukuran merupakan model yang mendeskripsikan hubungan antara variabel laten dengan variabel indikator. Model pengukuran terbagi menjadi dua bentuk yaitu model pengukuran variabel eksogen dan pengukuran variabel endogen. Model Pengukuran variabel eksogen dapat terjadi jika variabel indikator dipengaruhi oleh variabel laten, sedangkan model

pengukuran yaitu model dengan variabel endogen dipengaruhi oleh variabel laten. Persamaan untuk model pengukuran variabel eksogen adalah sebagai berikut:

$$X_1 = \lambda_{11}\xi_1 + \delta_1 \quad (4)$$
$$X_2 = \lambda_{21}\xi_1 + \delta_2 \quad (5)$$
$$X_3 = \lambda_{12}\xi_2 + \delta_3 \quad (6)$$
$$X_4 = \lambda_{22}\xi_2 + \delta_4 \quad (7)$$
$$X_5 = \lambda_{32}\xi_2 + \delta_5 \quad (8)$$

Sedangkan persamaan untuk model pengukuran variabel endogen :

$$Y_1 = \lambda_{13}\eta_1 + \varepsilon_1 \quad (9)$$
$$Y_2 = \lambda_{23}\eta_1 + \varepsilon_2 \quad (10)$$
$$Y_3 = \lambda_{33}\eta_1 + \varepsilon_3 \quad (11)$$
$$Y_4 = \lambda_{14}\eta_2 + \varepsilon_4 \quad (12)$$
$$Y_5 = \lambda_{24}\eta_2 + \varepsilon_5 \quad (13)$$
$$Y_6 = \lambda_{15}\eta_3 + \varepsilon_6 \quad (14)$$
$$Y_7 = \lambda_{25}\eta_3 + \varepsilon_7 \quad (15)$$

2.2 Confirmatory Factor Analysis (CFA)

Menurut [5] *Confirmatory Factor Analysis* (CFA) adalah suatu metode yang dapat mengevaluasi suatu model konstruk yang terdiri atas pengujian validitas dan reliabilitas. Pengujian validitas dilakukan untuk melihat seberapa mampu variabel konstruk (indikator) menjelaskan variabel latennya melalui nilai *loading factor*-nya. Nilai cut of value yang dijadikan pembanding adalah minimal *loading* > 0,30 dan pengujian reliabilitas diukur dengan membanding nilai *composite reliability* (CR) dengan nilai cut of value yakni dikatakan reliabel jika nilai CR ≥ 0,70 [6].

2.3 Uji kecocokan model

Evaluasi kecocokan model terhadap data terdiri dari beberapa tahapan, yaitu yang pertama adalah penentuan model yang baik untuk data di dalam SEM, hal ini dilakukan dengan pengujian semua uji statistik *Goodness of Fit* (GOF). Model yang baik mempunyai selisih yang kecil antara data ril dan data dugaan. Hasil dugaan yang memberikan galat yang tinggi, hasilnya tidak akan baik [7]. Pada metode SEM tidak terdapat statistik uji terbaik yang dapat memprediksi kecocokan model dengan data [6].

Selanjutnya adalah kecocokan model struktural. Tahapan ini dilakukan untuk memastikan beberapa hubungan yang dihipotesiskan pada model konseptualisasi didukung oleh data.

3. Metode Penelitian

Data pada penelitian ini adalah data yang bersumber dari survei status gizi remaja di Kabupaten Aceh Besar pada tahun 2018 yang dilakukan oleh Laboratorium Biostatistika Jurusan Statistika FMIPA Universitas Syiah Kuala. Variabel yang dibentuk adalah 1 variabel laten endogen dan 4 variabel laten eksogen (bebas). Variabel laten endogen berupa status gizi remaja sedangkan variabel laten eksogen berupa kebiasaan dan pola makan, status dan kondisi kesehatan, kondisi keluarga, dan pengetahuan remaja.

Tabel 1 Variabel pada penelitian

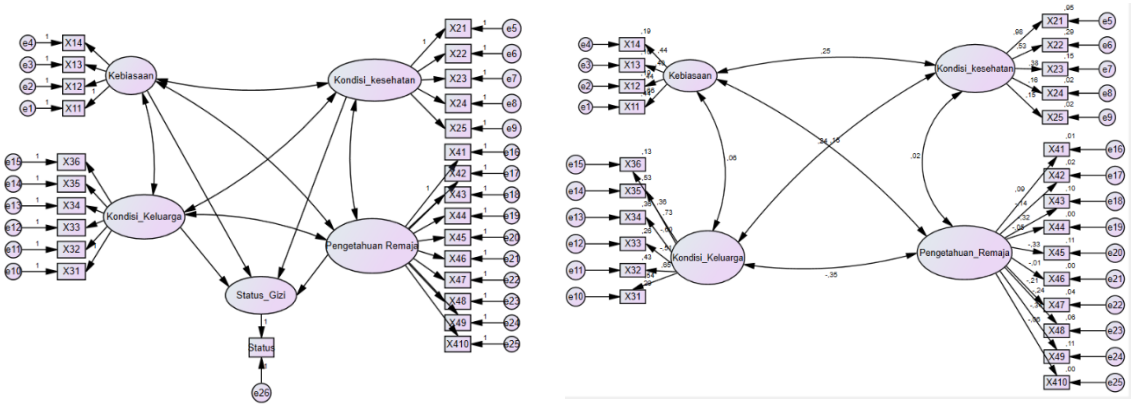
Variabel	Simbol	Variabel Indikator	Ket
Status gizi	*Y*	Kesehatan	Jumlah
		Frekuensi makan dalam sehari (X_{11})	Jumlah

Variabel	Simbol	Variabel Indikator	Ket
Kebiasan dan pola makan	x_1	Frekuensi sarapan dalam satu minggu terakhir (X_{12})	Hari
		Frekuensi makan siang dalam satu minggu terakhir (X_{13})	Hari
		Frekuensi makan malam dalam satu minggu terakhir (X_{14})	Hari
Status dan kondisi kesehatan	x_2	Tinggi Badan (X_{21})	Cm
		Berat Badan (X_{22})	Kg
		Umur (X_{23})	Tahun
		Kadar Hb (X_{24})	gr/dL
		Tekanan Darah (X_{25})	mmHg
Kondisi keluarga	x_3	Pendidikan terakhir ayah/wali (X_{31})	Tingkat Pendidikan
		Pendidikan terakhir ibu/wali (X_{32})	
		Pekerjaan terakhir ayah/wali (X_{33})	Jenis pekerjaan
		Pekerjaan terakhir ibu/wali (X_{34})	
		Jumlah pendapatan orang tua ayah/wali (X_{35})	Rupiah
		Uang saku perbulan (X_{36})	Rupiah
Pengetahuan remaja	x_4	Pengetahuan gizi (X_{41})	Skor
		Pengetahuan beresiko anemia (X_{42})	
		Pengetahuan ciri-ciri anemia (X_{43})	
		Pengetahuan penyebab anemia (X_{44})	
		Pengetahuan penyebab obesitas (X_{45})	
		Pengetahuan frekuensi makan (X_{46})	
		Pengetahuan pentingnya sarapan (X_{47})	
		Pengetahuan pentingnya makan siang (X_{48})	
		Pengetahuan pentingnya makan malam (X_{49})	
		Pengetahuan pentingnya olahraga (X_{410})	

Adapun tahapan analisis sebagai berikut. Pada tahap awal dilakukan pengembangan model teoritis. Pengunaan SEM adalah tidak untuk menghasilkan sebuah model, melainkan untuk mengkonfirmasi model teoritis tersebut melalui data empirik. Model teoritis yang telah dibentuk pada tahap pertama akan digambarkan dalam sebuah diagram jalur, yang akan mempermudah melihat hubungan kausalitas yang akan diuji. Kemudian mengkonversi diagram jalur kedalam persamaan struktural dan spesifikasi model pengukuran. Tahapan pengujian asumsi dilakukan untuk melihat seberapa besar ketepatan model yang digunakan dengan melakukan pengujian ukuran sampel, normalitas data, *outliers*, multikolinearitas dan singulatiras. Uji kecocokan model merupakan tahap akhir sebelum dilakukan interpretasi model.

4. Hasil dan Pembahasan

Ukuran sampel yang digunakan yaitu sebesar 234 sampel yang merupakan hasil terbaik dari eliminasi dari 242 data awal yang mana terdapat 8 data yang terindikasi menggunakan *Mahalanobis d-squared* pada output AMOS berdistribusi tidak normal sehingga perlu dilakukan elimiasi jumlah data sampel untuk mendapatkan hasil yang maksimal. Evaluasi ukuran sampel yang digunakan telah memenuhi minimum ukuran sampel dengan perbandingan 10 observasi pada 22 variabel teramati yaitu sebesar 220 sampel minimal. Penggunaan sampel sebanyak 234 didasarkan pada penggunaan metode estimasi *Maximum Likelihood* yang digunakan di mana metode sangat cocok untuk jumlah sampel 100 hingga 250 dan dalam hal ini asumsi distribusi kenormalan data merupakan bagian dari pengujian SEM menggunakan AMOS.



Gambar 1 (a) Diagram jalur; (b) *Standardized Loading Factor*.

Gambar 1(a) menunjukkan diagram jalur untuk variabel yang digunakan dalam penelitian. Sedangkan Gambar 1(b) memperlihatkan bahwa tidak semua variabel konstruk dapat menjelaskan variabel latennya. Terdapat nilai *Standardized Loading Factor* pada masing-masing kontruk < 0,30 atau dapat dikatakan tidak valid sehingga perlu dilakukan eliminasi pada variabel konstruk yang tidak sesuai ketentuan.

Tabel 2 Pengujian Reliabilitas

Variabel Laten	CR	*Cut of Value*	Keterangan
Kebiasan dan Pola_Makan	0,261	> 0,70	Tidak Reliabel
Status dan kondisi_kesehatan	0,7	> 0,70	Reliabel
Kondisi Keluarga	0,005	> 0,70	Tidak Reliabel
Pengetahuan Remaja	0,615	> 0,70	Tidak Reliabel
Status Gizi Remaja	0,254	> 0,70	Tidak Reliabel

Tabel 2 menjelaskan bahwa hanya variabel laten status dan kondisi kesehatan yang reliabel sedangkan variabel laten lainnya yakni kebiasaan dan pola makan, kondisi keluarga, pengetahuan remaja, dan status gizi remaja tidak reliabel. Meskipun demikian data tetap dapat dilanjutkan menggunakan pengujian SEM karena data bersifat variasi bukan menggunakan data skala likert.

4.1 Kecocokan model keseluruhan

Tahapan pengujian kecocokan model kesuluruhan berkaitan dengan analisis ukuran kecocokan statistik yang terdapat pada *output* AMOS.

Tabel 3 Kecocokan model

Ukuran Kecocokan	Target Tingkat Kecocokan	Estimasi	Tingkat Kecocokan
Statistic Chi-Square (X^2)	X^2 Nilai kecil	69,364	Baik
	P-value > 0,05	0,108	Baik
	Normed Chi-Square (CMIN) < 2	1,239	Baik
Root Mean Square Error of Approximation (RMSEA)	RMSEA < 0,05 dinyatakan*close fit*	0,032	Baik

Ukuran Kecocokan	Target Tingkat Kecocokan	Estimasi	Tingkat Kecocokan
Goodness-of-Fit Index (GFI)	GFI ≥ 0,9 dinyatakan baik, jika nilai 0,8 ≤ GFI ≥ 0,9 maka dinyatakan kurang baik	0,956	Baik
Comparative Fit Index (CFI)		0,949	Baik
Adjusted Goodness of Fit Index (AGFI)	≥ 0,9 dinyatakan baik, jika nilai 0,8−0,9 maka dinyatakan kurang baik	0,928	Baik
Tucker-Lewis Index atau *Non-Normed Fit Index* (TLI atau NNFI)		0,929	Baik

Pada Tabel 3 bahwa hasil pengukuran yang ditampilkan adalah ukuran kecocokan model yang baik. Hal ini mengasumsikan bahwa model *structural* secara keseluruhan baik karena telah dilakukan eliminasi konstruk pada variabel laten setelah dilakukan pengujian *Confirmatory Factor Analysis* (CFA). Hasil tersebut memungkinkan hasil maksimal pada pengujian menggunakan analisis jalur SEM.

4.2 Pengujian Asumsi SEM

Pengujian asumsi SEM bertujuan untuk melihat ketentuan-ketentuan pengujian asumsi SEM sebelum dilakukan pengujian SEM. Pengujian *outlier* secara *univariate* ditentukan dengan melihat nilai *Zscore* pada *output* SPSS, dengan membandingkan nilai *output* > 3 maka terdapat outlier pada data penelitian. Terdapat nilai *Zscore* > 3 yakni pada data *Zscore* X_{13} dan X_{22}. Meskipun demikian, secara keseluruhan data yang digunakan dapat dilakukan pengujian menggunakan analisis SEM.

Pengujian *outlier* secara *multivariate* dapat dilakukan dengan memantau nilai P_1 dan P_2 pada *output Mahalanobis d-squared*, yakni tidak terjadi outlier jika P_1 dan P_2 > 0,05. Hasil yang diperoleh adalah terdapat nilai P_1 > 0,05 yaitu pada observasi nomor 190 dan 115. Namun pada P_2 banyak nilai P_2 > 0,05 yaitu observasi nomor 134, 87, 58, 136, 139, 101, 202, 100, 168, 152, 190, 115. Hal ini menyimpulkan bahwa secara univariate dan multivariate data dapat dilanjukan pengujian menggunakan model analisis SEM.

Selanjutnya, pengujian normalitas data dilakukan secara univariate dan multivariate dengan membandingkan nilai c.r dengan 2,58. Data dinyatakan berdistribusi normal jika nilai c.r < 2,58. Pengujian multivariate dapat menjelaskan bahwa nilai c.r sebesar 1,796 yakni < 2,58. Dapat dinyatakan bahwa data berdistibusi normal.

Pengujian evaluasi nilai residual dapat dilakukan dengan membandingkan nilai residual dengan 2,58. Terdapat nilai residu pada data jika nilai residu > 2,58. Berdasarkan *output standardized residual*, dapat disimpulkan bahwa tidak terdapat nilai residual yang tinggi pada data penelitian dengan melihat nilai residual yang berada dibawah 2,58.

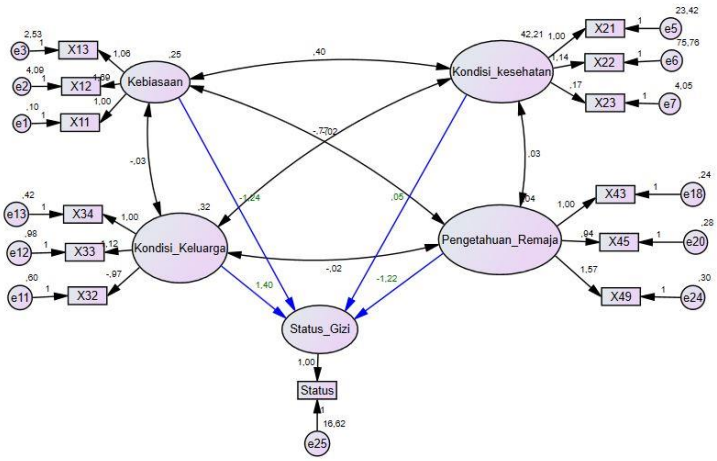
Pengujian *Multicollinierity* dan *Singularity* ditentukan berdasarkan *Determinant of sample covariance matrix*. *Multicollinierity* dan *Singularity* terjadi jika nilai *Determinant of sample covariance matrix* menjauhi angka nol (0). Berdasarkan output AMOS, didapat hasil bahwa nilai *determinant of sample covariance matrix* = 26876,231. Hal ini dapat menjelaskan bahwa tidak terjadi *Multicollinierity* dan *Singularity* pada data.

Berdasarkan pengujian asumsi SEM, dapat dikatakan bahwa data lolos dari pengujian asumsi SEM dan dapat dilakukan pengujian menggunakan model analisis jalur SEM.

Journal of Data Analysis
Vol.2, No.2, December 2019, p. 88 - 95

4.3 Pengembangan Model

Model dikembangkan dengan menggambarkannya ke bentuk diagram jalur. Diagram jalur yang dibentuk dapat digambarkan sebagai berikut.



Gambar 2 Diagram jalur dari model.

Gambar 2 menampilkan hubungan dari indikator-indikator terhadap variabel latennya dan variabel laten eksogen yang memengaruhi variabel laten endogen. Nilai yang berada di antara variabel laten dengan indikator adalah *loading factor* untuk model pengukuran. Sedangkan nilai yang berada pada garis yang menghubungkan antar variabel laten merupakan *loading factor* untuk model structural.

Setelah model dikembangkan ke dalam diagram jalur, kemudian konversi ke dalam persamaan pengukuran dan *structural*. Persaman model pengukuran dan *structural* yang didapatkan sebagai berikut. Adapun model pengukuran yang dapat dibentuk adalah :

X_{11} = 1,000 Kebiasaan + 0,100
X_{12} = 1,686 Kebiasaan + 4,089
X_{13} = 1,057 Kebiasaan + 2,529

X_{32} = -0,966 Kondisi Keluarga + 0,596
X_{33} = 1,122 Kondisi Keluarga + 0,982
X_{34} = 1,000 Kondisi Keluarga + 0,420

X_{21} = 1,000 Kondisi Kesehatan + 23,419
X_{22} = 1,137 Kondisi Kesehatan + 75,763
X_{23} = 0,168 Kondisi Kesehatan + 4,055

X_{43} = 1,000 Pengetahuan Remaja + 0,240
X_{45} = 0,939 Pengetahuan Remaja + 0,283
X_{49} = 1,000 Pengetahuan Remaja + 0,304

Status (Y) = 1,000 + Status Gizi + 16,618

Tabel 4 Model Struktural

			Estimate	**C.R.**	***P-Value***
Status Gizi Remaja	<-	Kebiasan dan Pola Makan	-1,238	-1,661	0,097
Status Gizi Remaja	<-	Status dan kondisi kesehatan	0,054	1,041	0,298
Status Gizi Remaja	<-	Kondisi Keluarga	1,397	2,004	0,045
Status Gizi Remaja	<-	Pengetahuan Remaja	-1,217	-0,453	0,651

Berdasarkan Tabel 4 dapat dibentuk persamaan *structural* sebagai berikut:

Status gizi remaja = −1,238 Kebiasaan + 0,054 kondisi Kesehatan + 1,397 Kondisi Keluarga −1,217 Pengetahuan Remaja

Berdasarkan model *structural* yang diperoleh, dapat kita lihat bahwa variabel-variabel yang mempengaruhi status gizi remaja yang bernilai positif adalah kondisi kesehatan dan kondisi keluarga. Sedangkan kebiasaan dan pola makan dan pengetahuan remaja bernilai negatif terhadap status gizi remaja. Jika dilihat dari pengaruh signifikansinya hanya kondisi keluarga yang berpengaruh dengan positif dan signifikan terhadap status gizi remaja. Hal ini terlihat dari nilai *P-value* pada kondisi keluarga yaitu 0,045 < 0,05 dan 0,1. Artinya, kondisi keluarga berpengaruh positif signifikan pada taraf kepercayaan 90% hingga 95% terhadap status gizi remaja di kabupaten Aceh Besar. Dengan demikian peran keluarga dan kondisi kesehatan remaja dapat mendorong peningkatan status gizi remaja. Hal ini menjadi perhatian penting bagi keluarga dan lembaga lainnya yang berkecimpung di bidang kesehatan untuk dapat melakukan penyuluhan dan pengarahan agar status gizi remaja terkendali dengan baik.

5 Kesimpulan

Kesimpulan yang dapat diambil berdasarkan hasil penelitian adalah terbentuknya dugaan *Structural Equation Modeling* (SEM) terhadap variabel-variabel yang berkaitan dengan status gizi remaja di Kabupaten Aceh Besar sebagai beriku :

Status gizi remaja = −1,238 Kebiasaan + 0,054 kondisi Kesehatan + 1,397 Kondisi Keluarga −1,217 Pengetahuan Remaja

Beberapa faktor yang dapat memengaruhi status gizi remaja di Kabupaten Aceh Besar adalah status dan kondisi kesehatan serta kondisi keluarga di mana nilai parameter bernilai positif. Sedangkan kebiasaan dan pola makan dan pengetahuan remaja bernilai negatif. Hal ini menunjukkan bahwa kondisi keluarga berpengaruh secara langsung dan signifikan terhadap status gizi remaja di Kabupaten Aceh Besar.

Daftar Pustaka

[1] Santoso, S. 2010. *Statistik Multivariat Konsep dan Aplikasi SPSS*. PT. Elex Media Komputindo, Jakarta.
[2] Azyus, A. N. 2016. *Hubungan Konsumsi Fast Food dengan Status Gizi pada Pelajar SMA Banda Aceh dan Aceh Besar*. Syiah Kuala University, Banda Aceh.
[3] Kesuma, Z. M. dan Rahayu, L. 2017. Identifikasi status gizi pada remaja di kota Banda Aceh. Statistika, vol. 17, no. 2, pp. 63–69.
[4] Sujarweni, V. W. 2015. *Statistik untuk Bisnis dan Ekonomi*. Pustaka Baru Press, Yogyakarta.
[5] Bahri, S. and Zamzam, F. 2015. *Model Penelitian Kuantitatif Berbasis SEM-AMOS*. CV. Budi Utama, Yogyakarta.
[6] Hair, J.F., dan Black, W.C., Babin, B.J., and Anderson, R.E. 1998. *Multivariate Data Analysis*. McGraw-Hill, New Jersy.
[7] Marzuki, Sofyan, H. dan Rusyana A. 2010. Pendugaan Selang Kepercayaan Persentil Bootstrap Nonparametrik untuk Parameter Regresi. Statistika, Vol 10 No.1, pp. 13-23: 2599-2538.

Journal of Data Analysis
Vol.2, No.2, December 2019, p. 96-108

Pemodelan Panel Spasial terhadap Faktor-Faktor yang Memengaruhi Kesehatan di Provinsi Papua

Ira Rosianal Hikmah[1*], Yulial Hikmah[2]

[1]Sekolah Tinggi Sandi Negara, Bogor, Jawa Barat, Indonesia
[2]Program Pendidikan Vokasi Universitas Indonesia, Depok, Jawa Barat, Indonesia
Email: ira.rosianal@stsn-nci.ac.id[*], yulialhikmah47@ui.ac.id

* = *corresponding author*

Abstrak

Menurut data BPS, tingkat pertumbuhan populasi penduduk di Indonesia secara konsisten meningkat setiap tahun. Kondisi pertumbuhan populasi penduduk yang tidak dapat ditekan akan menyebabkan berbagai masalah. Salah satu masalah yang mungkin terjadi di Indonesia dan sulit diselesaikan adalah kesehatan masyarakat Indonesia. Provinsi Papua menjadi provinsi dengan persentase rumah tangga kumuh perkotaan tertinggi di Indonesia, persentase rumah tangga yang memiliki akses terhadap layanan sanitasi layak dan berkelanjutan terendah, menempati peringkat kelima persentase terendah yang memiliki akses terhadap layanan sumber air minum layak dan berkelanjutan, serta menjadi provinsi dengan angka kematian balita per 1000 kelahiran hidup tertinggi di Indonesia. Salah satu penyebabnya adalah rendahnya persentase balita yang pernah mendapatkan imunisasi. Penelitian ini melakukan pemodelan untuk mendapatkan faktor-faktor yang memengaruhi kesehatan di Provinsi Papua. Penelitian ini melibatkan pengaruh spasial (model panel spasial) dan membandingkannya dengan model panel biasa untuk mendapatkan model terbaik. Model panel spasial yang dipilih dalam penelitian ini adalah model SAR, SEM, dan GSM. Hasil menunjukkan bahwa model SAR dengan pengaruh tetap adalah model terbaik dalam penelitian ini.

Informasi Artikel

Sejarah Artikel:
Diajukan 3 Mei 2020
Diterima 26 Mei 2020

Kata Kunci :
Kesehatan Masyarakat
Provinsi Papua
Model Panel Spasial
SAR
SEM
GSM

Abstract

According to BPS data, the rate of population growth in Indonesia consistently increasing every year. Conditions of population growth that cannot be suppressed will cause several problems. One of them and difficult to solve is the public health problem. Papua Province is the province with the highest percentage of urban slum households, the lowest percentage of households that has access to decent and sustainable sanitation services, ranks fifth lowest who has access to decent and sustainable drinking water services, and the highest number infant mortality per 1000 live births in Indonesia. One of the reasons is the low percentage of children under five who have been immunized. This research is modeling to find the factors that influence health in Papua Province. This research involves spatial influence (spatial panel model) and compares it with the

Keywords:
Public Health
Papua Province
Spatial Panel Model
SAR
SEM
GSM

ordinary panel models to get the best model. The spatial panel models selected in this research are SAR, SEM, and GSM models. The results show that the SAR model with the fixed effect is the best in this research.

1. Pendahuluan

Badan Pusat Statistik atau BPS, pada tahun 2010, menunjukkan bahwa laju pertumbuhan penduduk di Indonesia meningkat sebesar 1,49% pertahunnya [1]. Kondisi pertumbuhan penduduk yang tidak dapat ditekan akan menimbulkan beberapa permasalahan. Salah satu permasalahan di Indonesia dan sulit untuk diselesaikan adalah masalah kesehatan masyarakat. Hal ini dikarenakan sebagian masyarakat Indonesia belum mendapatkan pelayanan kesehatan karena biaya kesehatan yang terus meningkat. Permasalahan kesehatan menjadi salah satu fokus pemerintah selain masalah kemiskinan dan pendidikan. Sejumlah program dan kebijakan telah dilaksanakan untuk meningkatkan tingkat kesehatan masyarakat. Salah satunya adalah Program Indonesia Sehat yang digalakkan sejak tahun 2014 [2]. Akan tetapi, hingga kini permasalahan kesehatan di Indonesia masih cukup tinggi.

Riset Kesehatan Dasar (Riskesdas) merupakan riset berskala nasional yang dilakukan oleh Badan Penelitian dan Pengembangan Kesehatan (Litbangkes) Kementerian Kesehatan RI untuk mendapatkan gambaran mengenai bagaimana kesehatan masyarakat Indonesia. Riskesdas pertama kali dilakukan pada tahun 2007, kemudian dilakukan lagi tahun 2010, 2013, dan yang terbaru adalah tahun 2018. Data riskesdas meliputi kondisi kesehatan masyarakat dari berbagai usia, jenis kelamin, penyakit, serta gaya hidup [3]. Hasil utama Riskesdas 2018 menyebutkan bahwa provinsi dengan prevalensi ispa dan pneumonia tertinggi di Indonesia adalah Provinsi Papua. Sedangkan untuk prevalensi TB paru, Provinsi Papua tertinggi kedua setelah Banten. Selain itu, Provinsi Papua juga menjadi daerah prevalensi diare pada balita tertinggi di Indonesia [4]. Oleh karena itu, penelitian ini memfokuskan pada permasalahan kesehatan masyarakat di Provinsi Papua.

Terdapat banyak faktor yang memengaruhi permasalahan kesehatan di Provinsi Papua, di antaranya adalah faktor lingkungan dan gaya hidup [4]. Menurut data BPS, pada tahun 2018, Provinsi Papua menjadi provinsi dengan persentase rumah tangga kumuh perkotaan tertinggi di Indonesia [5], persentase rumah tangga yang memiliki akses terhadap layanan sanitasi layak dan berkelanjutan terendah [6], menempati peringkat kelima persentase terendah yang memiliki akses terhadap layanan sumber air minum layak dan berkelanjutan [7], dan menjadi provinsi dengan angka kematian balita per 1000 kelahiran hidup tertinggi [8]. Salah satu penyebabnya adalah rendahnya persentase balita yang pernah mendapatkan imunisasi sehingga mudah terkena penyakit. Selain itu, pendidikan mengenai gaya hidup sehat di Provinsi Papua juga cukup rendah. Salah satunya terlihat dari hasil utama Riskesdas 2018 bahwa penduduk usia lebih dari 10 tahun di Provinsi Papua yang mencuci tangan dengan benar merupakan yang terendah kedua setelah NTT [4]. Oleh karena itu, fokus penelitian adalah pada faktor-faktor tersebut, yaitu persentase keluhan kesehatan, rata-rata lama sakit, ketersediaan fasilitas BAB, ketersediaan sumber air minum, dan persentase imunisasi balita.

Model panel spasial merupakan salah satu pengembangkan model data panel dengan mempertimbangkan dependensi dari wilayah atau lokasi. Hal tersebut dapat menentukan korelasi suatu lokasi dengan lokasi lain yang berdekatan. Beberapa penelitian telah menggunakan faktor

spasial dalam permasalahan kependudukan seperti pada penelitian Anggraeni [9] dan Hikmah [10]. Oleh karena itu, penelitian ini akan melibatkan pengaruh spasial atau lokasi untuk memodelkan dan menentukan faktor-faktor yang memengaruhi kesehatan di Provinsi Papua.

2. Landasan Teori

2.1. Model Data Panel

Terdapat tiga model dalam data panel yaitu model pengaruh tetap, pengaruh acak, dan gabungan. Persamaan model panel pengaruh tetap adalah sebagai berikut:

$$y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{jit} + \mu_i + \varepsilon_{it}$$

Sedangkan persamaan model panel pengaruh acak adalah:

$$y_{it} = \beta_{0i} + \sum_{j=1}^{k} \beta_j x_{jit} + \varepsilon_{it}$$

dengan, $\beta_{0i} = \beta_0 + \mu_i$. Pada model panel pengaruh gabungan, persamaan modelnya sama dengan regresi linier biasa yaitu:

$$y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{jit} + \varepsilon_{it}$$

dengan $i = 1, \dots, N$, $t = 1, \dots, T$, N menyatakan banyaknya individu, T menyatakan banyaknya periode waktu, dan k menyatakan banyaknya variabel penjelas.
[11].

2.2. Uji Breusch-Pagan

Pangestika [12] melakukan pengujian apakah ada efek waktu, individu atau keduanya dengan menggunakan Uji Breusch-Pagan. Hipotesis nolnya adalah:

$H_0^{(1)}$: Tidak ada efek individu dan waktu

$H_0^{(2)}$: Tidak ada efek individu

$H_0^{(3)}$: Tidak ada efek waktu

2.3. Uji Chow dan Uji Hausman

Pengujian untuk menentukan model yang digunakan, baik model acak, tetap, atau gabungan, dapat dilakukan dengan menggunakan Uji Chow dan Uji Hausman. Uji Chow digunakan untuk menguji signifikansi antara model gabungan dan model pengaruh tetap sedangkan uji Hausman digunakan untuk menguji signifikansi antara model pengaruh acak dan model pengaruh tetap. Statistik uji yang digunakan pada Uji Chow adalah:

$$F = \frac{(RRSS - URSS)/(N-1)}{URSS/(NT - N - k)}$$

dengan $RRSS$ $(Restricted\ Residual\ Sums\ of\ Square)$ merupakan jumlah kuadrat eror hasil pendugaan model gabungan dan $URSS$ $(Unrestricted\ Residual\ Sums\ of\ Square)$ merupakan jumlah kuadrat eror hasil pendugaan model pengaruh tetap. Sedangkan statistik uji pada Uji Hausman adalah:

$$\chi^2 = \hat{q}[Var(\hat{q})]^{-1}\hat{q}$$

dengan $\hat{q} = \hat{\beta}_{acak} - \hat{\beta}_{tetap}$ [11].

2.4. Autokorelasi Spasial

Autokorelasi spasial merupakan ukuran korelasi antara suatu variabel dengan dirinya sendiri berdasarkan lokasi. Salah satu statistik yang umum digunakan adalah Indeks Moran Global (I) yaitu:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2}$$

dengan n menyatakan banyaknya lokasi, z_i dan z_j masing-masing menyatakan nilai pengamatan pada lokasi ke-i dan ke-j, $\bar{z}$ menyatakan rata-rata nilai pengamatan, dan w_{ij} merupakan pembobot yang diberikan antara lokasi ke-i dan ke-j.

Pengujian hipotesis untuk Indeks Moran Global adalah:

Hipotesis Uji:

$H_0 : I = 0$

$H_1^{(1)}: I > 0$

$H_1^{(2)}: I < 0$

Statistik Uji:

$Z(I) = \frac{I - E(I)}{\sigma(I)} \sim N(0,1)$ [13].

2.5. Model Panel Spasial

Anselin [13] menyatakan model panel dengan GSM, SEM, dan SAR sebagai berikut:

1. Model Panel GSM:

$$y_{it}^{*} = \rho \sum_{j=1}^{N} w_{ij} y_{it}^{*} + \beta_0 + \sum_{j=1}^{k} \beta_j x_{jit} + \mu_i + \phi_{it}$$

dengan $\phi_{it} = \lambda \sum_{j=1}^{N} w_{ij}\phi_{it} + \varepsilon_{it}$.

2. Model Panel SEM:

$$y_{it}^{*} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{jit} + \mu_i + \phi_{it}$$

dengan $\phi_{it} = \lambda \sum_{j=1}^{N} w_{ij}\,\phi_{it} + \varepsilon_{it}$

3. Model Panel SAR:

$$y_{it}^{*} = \rho \sum_{j=1}^{N} w_{ij}\, y_{it}^{*} + \beta_0 + \sum_{j=1}^{k} \beta_j x_{jit} + \mu_i + \varepsilon_{it}$$

3. Metodologi Penelitian

3.1 Data dan Variabel Penelitian

Data penelitian ini merupakan data sekunder yang bersumber dari data Badan Pusat Statistik (BPS). Data panel yang digunakan adalah 29 data kab/kota di Provinsi Papua dengan periode waktu 5 tahun yaitu mulai tahun 2013 hingga tahun 2017 dengan keterangan variabel sebagai berikut:

Tabel 1 Variabel Penelitian

Variabel	Satuan
$Persentase\ penduduk\ sakit\ (\boldsymbol{Y})$	%
$Persentase\ keluhan\ kesehatan\ (\boldsymbol{X1})$	%
$Rata-rata\ lama\ sakit\ (\boldsymbol{X2})$	Hari
$Persentase\ rumah\ tangga\ dengan\ fasilitas\ BAB\ sendiri\ (\boldsymbol{X3})$	%
$Persentase\ rumah\ tangga\ dengan\ fasilitas\ BAB\ bersama\ (\boldsymbol{X4})$	%
$Persentase\ balita\ tidak\ imunisasi\ BCG\ (\boldsymbol{X5})$	%
$Persentase\ balita\ tidak\ imunisasi\ DPT\ (\boldsymbol{X6})$	%
$Persentase\ balita\ tidak\ imunisasi\ POLIO\ (\boldsymbol{X7})$	%
$Persentase\ balita\ tidak\ imunisasi\ CAMPAK\ (\boldsymbol{X8})$	%
$Persentase\ balita\ tidak\ imunisasi\ HEPATITIS\ B\ (\boldsymbol{X9})$	%
$Persentase\ rumah\ tangga\ dengan\ sumber\ air\ minum\ sumur\ terlindungi\ (\boldsymbol{X10})$	%

3.2 Metode Penelitian

Berikut ini merupakan tahapan penelitian yang dilakukan:

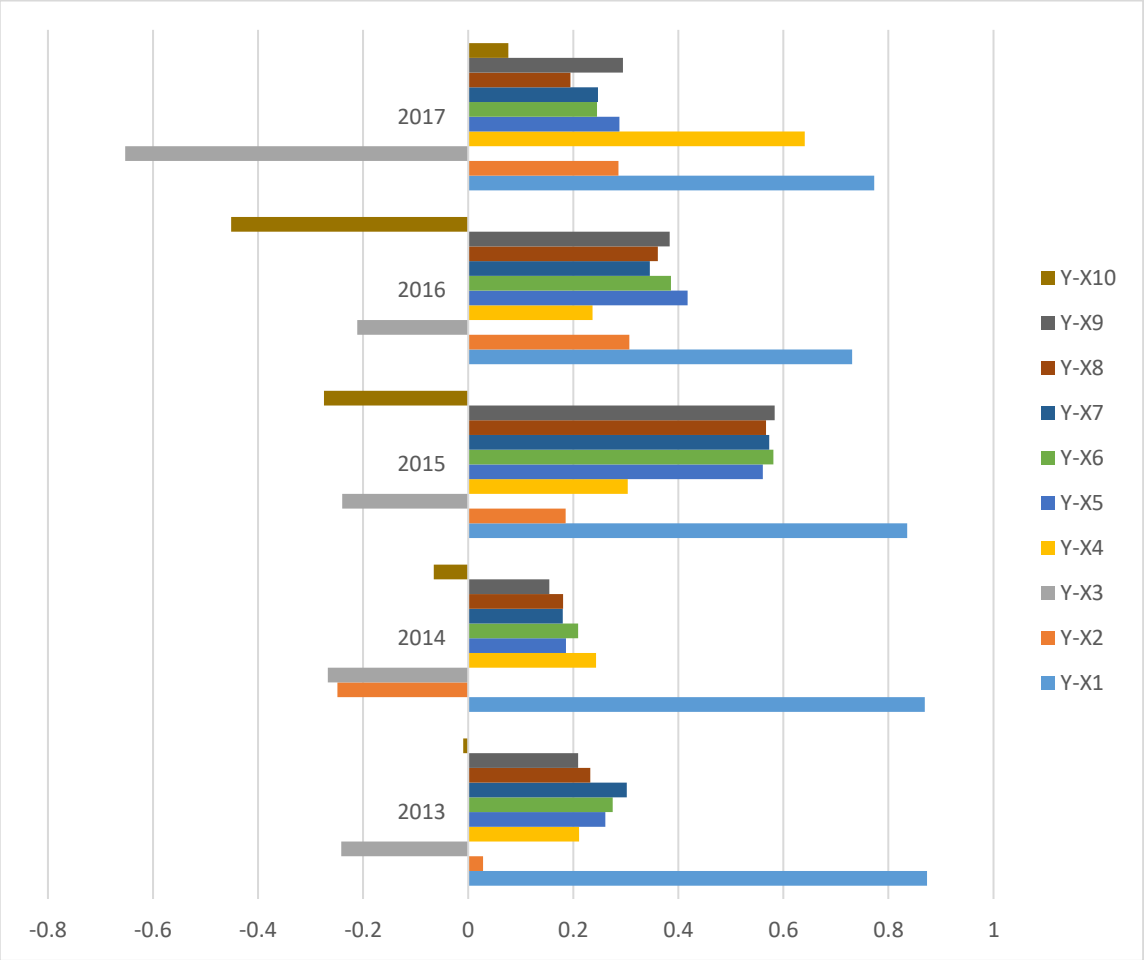
1. Menghitung korelasi antar variabel penelitian.
2. Menganalisis dengan model data panel.
3. Melihat ada atau tidaknya ketergantungan spasial dengan korelasi Indeks Moran Global.
4. Menganalisis dengan model panel spasial.
5. Memilih model terbaik dengan melihat nilai R^2 terbesar.

Penelitian ini memanfaatkan perangkat lunak Microsoft Excel dan R [14] dalam melakukan analisis deskriptif dan pemodelan.

4. Hasil dan Pembahasan

4.1. Menghitung Korelasi antar Variabel Penelitian

Gambar 1 menunjukkan korelasi antar variabel bebas dengan variabel terikatnya di setiap tahunnya. Berdasarkan Gambar 1 terlihat bahwa persentase keluhan kesehatan ($X1$) berkorelasi positif terhadap persentase penduduk sakit (Y). Selain itu, dapat dilihat pula bahwa persentase rumah tangga dengan fasilitas BAB sendiri ($X3$) memberikan korelasi negatif di setiap tahunnya. Variabel yang berpengaruh paling signifikan adalah variabel persentase keluhan kesehatan.



Gambar 1 Koefisien korelasi antar X dengan Y

4.2. Menganalisis dengan Model Data Panel

Pada tahapan penelitian ini, terlebih dahulu dilakukan uji Breusch-Pagan untuk melihat apakah ada efek waktu, individu, atau waktu dan individu secara bersamaan. Tabel 2 menunjukkan hasil uji Breusch-Pagan. Dengan menggunakan taraf signifikansi sebesar 5%, diperoleh hasil bahwa terdapat efek individu pada model data panel.

Tabel 2 Hasil Uji Breusch-Pagan

Efek	χ^2	$p-value$
Waktu dan Individu	3,938	0,139
Waktu	0,065	0,798
Individu	3,875	0,049

Karena hasil pada Tabel 2 menunjukkan bahwa data penelitian memiliki efek individu, maka selanjutkan dilakukan pendugaan parameter pada model data panel baik dengan pengaruh tetap, acak, maupun gabungan keduanya.

Tabel 3 Hasil Pendugaan Parameter Model Pengaruh Tetap

Variabel	Koefisien	$p-value$
$X1$	0,625	$< 0{,}000$
$X2$	0,358	0,028
$X3$	$-0{,}019$	0,039
$X4$	0,005	0,027
$X5$	$-0{,}009$	0,818
$X6$	0,082	0,039
$X7$	$-0{,}016$	0,727
$X8$	$-0{,}029$	0,359
$X9$	$-0{,}041$	0,209
$X10$	$-0{,}001$	0,038
R^2		0,7932

Tabel 4 Hasil Pendugaan Parameter Model Pengaruh Acak

Variabel	Koefisien	$p-value$
$X1$	0,534	$< 0{,}000$
$X2$	0,417	0,038
$X3$	0,006	0,639
$X4$	0,007	0,609
$X5$	$-0{,}066$	0,137
$X6$	0,074	0,088
$X7$	0,041	0,401
$X8$	$-0{,}011$	0,729
$X9$	$-0{,}064$	0,058
$X10$	0,008	0,539
R^2		0,5853

Tabel 5 Hasil Pendugaan Parameter Model Pengaruh Gabungan

Variabel	Koefisien	$p-value$
$X1$	0,647	$< 0{,}000$
$X2$	0,34	0,029
$X3$	$-0{,}024$	0,040
$X4$	0,001	0,918
$X5$	0,013	0,757
$X6$	0,084	0,059
$X7$	$-0{,}041$	0,409
$X8$	$-0{,}034$	0,288
$X9$	$-0{,}036$	0,270
$X10$	$-0{,}003$	0,789
R^2		0,7326

Berdasarkan Tabel 3, Tabel 4, dan Tabel 5, terlihat bahwa pada model pengaruh tetap, variabel yang signifikan adalah variabel $X1, X2, X3, X4, X6$, dan $X10$. Pada model pengaruh acak, hanya variabel $X1$ dan $X2$ yang signifikan sedangkan untuk model pengaruh gabungan, hanya variabel $X1, X2$, dan $X3$ yang signifikan. Selanjutnya Tabel 6, 7, dan 8 menunjukkan hasil pendugaan parameter model yang hanya melibatkan variabel-variabel yang signifikan.

Tabel 6 Hasil Pendugaan Parameter Model Pengaruh Tetap Tanpa Melibatkan $X5$, $X7$, $X8$, dan $X9$

Variabel	Koefisien	$p-value$
$X1$	0,601	$< 0,000$
$X2$	0,416	0,009
$X3$	$-0,021$	0,020
$X4$	0,007	0,010
$X6$	0,084	0,027
$X10$	$-0,003$	0,029
R^2		0,7495

Tabel 7 Hasil Pendugaan Parameter Model Pengaruh Acak Melibatkan $X1$ dan $X2$

Variabel	Koefisien	$p-value$
$X1$	0,531	$< 0,000$
$X2$	0,469	0,038
R^2		0,5386

Tabel 8 Hasil Pendugaan Parameter Model Pengaruh Gabungan Melibatkan $X1, X2$, dan $X3$

Variabel	Koefisien	$p-value$
$X1$	0,629	$< 0,000$
$X2$	0,332	0,019
$X3$	$-0,022$	0,030
R^2		0,7159

Berdasarkan Tabel 6, 7, dan 8, penelitian ini memilih model pengaruh tetap karena menghasilkan R^2 yang paling besar. Selanjutnya dilakukan Uji Chow dan Uji Hausman.

Tabel 9 Hasil Uji Chow

F	$2,262$
$Db1$	26
$Db2$	115
$p-value$	0,002

Tabel 10 Hasil Uji Hausman

χ^2	$6,710$
Db	1
$p-value$	0,008

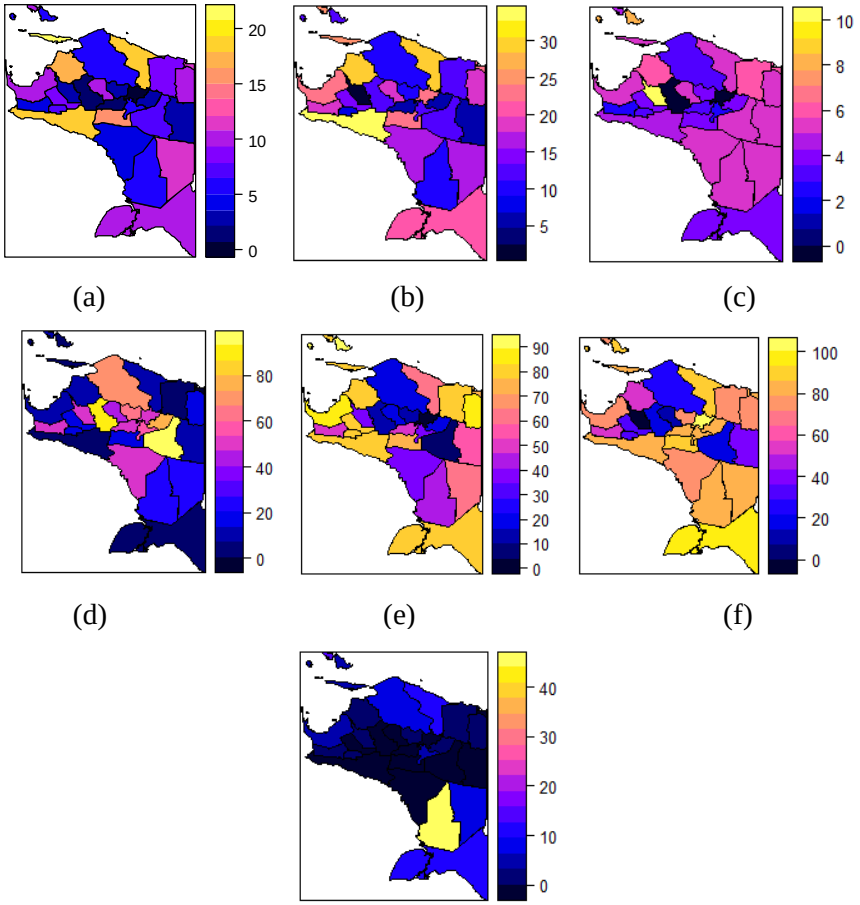
Berdasarkan Tabel 9 dan Tabel 10, maka diketahui bahwa model penelitian ini mengikuti model pengaruh tetap. Oleh karena itu, diperoleh model regresi data panel dengan pengaruh tetap sebagai berikut:

$y_{it} = 0,601\ x_{1it} + 0,416x_{2it} - 0,021\ x_{3it} + 0,007x_{4it} + 0,084x_{6it} - 0,003x_{10it} + \hat{\mu}_i + \varepsilon_{it}$

dengan R^2 = 0,7495, artinya sebesar hampir 75% variabel-variabel bebas pada model mampu menjelaskan variasi dari variabel terikatnya.

4.3. Menentukan Ketergantungan Spasial

Peta persebaran persentase penduduk sakit, persentase keluhan kesehatan, rata-rata lama sakit, persentase rumah tangga dengan fasilitas BAB sendiri, persentase rumah tangga dengan fasilitas BAB bersama, persentase balita tidak imunisasi DPT, dan sumber air minum sumur terlindungi di Provinsi Papua pada tahun 2017 terlihat pada gambar berikut:



Gambar 2 Peta Provinsi Papua Tahun 2017 (a) Y (b) $X1$ (c) $X2$ (d) $X3$ (e) $X4$ (f) $X6$ (g) $X10$

Pada peta di atas terlihat bahwa secara geografis terdapat kemiripan nilai pengamatan dengan tetangga kota/kab terdekatnya. Hal ini dapat dilihat pada warna peta yang menghasilkan warna yang identik yang mengindikasikan adanya dependensi spasial. Oleh karena itu, perlu dilakukan analisis data panel spasial. Dalam melakukan analisis data spasial, terlebih dahulu diperiksa apakah antar kota/kab saling bebas atau tidak dengan melakukan uji autokorelasi spasial. Hasil uji autokorelasi spasial dapat dilihat pada Tabel 11. Terlihat bahwa, setiap variabel, antar kab/kota saling dependen (tidak saling bebas). Oleh karena itu, selanjutnya dilakukan analisis panel spasial.

Tabel 11 Hasil Uji Autokorelasi Spasial

	$Z(I)$	$p-value$	$Keterangan$
Y	2,0653	0,042	$H_1: I > 0;\ Terdapat\ otokorelasi\ spasial\ positif$
$X1$	2,7863	0,008	$H_1: I > 0;\ Terdapat\ otokorelasi\ spasial\ positif$
$X2$	4,3524	0,033	$H_1: I > 0;\ Terdapat\ otokorelasi\ spasial\ positif$
$X3$	2,8422	0,002	$H_1: I > 0;\ Terdapat\ otokorelasi\ spasial\ positif$
$X4$	4,1936	$< 0{,}000$	$H_1: I > 0;\ Terdapat\ otokorelasi\ spasial\ positif$
$X6$	3,2374	0,000	$H_1: I > 0;\ Terdapat\ otokorelasi\ spasial\ positif$
$X10$	1,2987	0,009	$H_1: I > 0;\ Terdapat\ otokorelasi\ spasial\ positif$

4.4. Menganalisis dengan Model Panel Spasial

Berdasarkan hasil sebelumnya diperoleh bahwa data penelitian memiliki otokorelasi spasial. Hal ini menunjukkan bahwa mempertimbangan ketergantungan spasial merupakan pilihan yang tepat. Selanjutkan dilakukan pemodelan data panel spasial dan hasil pendugaan parameternya dapat dilihat pada Tabel 12, 13, dan 14 untuk model panel GSM, SAR, dan SEM.

Tabel 12 Pendugaan Parameter Model GSM dengan Pengaruh Tetap

Variabel	Koefisien	$p-value$
Rho	0,151	0,032
$Lambda$	$-0{,}669$	0,019
$X1$	0,539	$< 0{,}000$
$X2$	0,437	0,041
$X3$	0,006	0,569
$X4$	0,011	0,026
$X5$	$-0{,}078$	0,033
$X6$	0,050	0,158
$X7$	0,068	0,048
$X8$	$-0{,}014$	0,615
$X9$	$-0{,}059$	0,039
$X10$	0,004	0,712
R^2	0,8294	

Tabel 13 Pendugaan Parameter Model SEM dengan Pengaruh Tetap

Variabel	Koefisien	$p-value$
$Lambda$	$-0{,}446$	0,017
$X1$	0,541	$< 0{,}000$
$X2$	0,398	0,046
$X3$	0,011	0,288
$X4$	0,011	0,329
$X5$	$-0{,}080$	0,031
$X6$	0,061	0,028
$X7$	0,062	0,017
$X8$	$-0{,}009$	0,734
$X9$	$-0{,}064$	0,223
$X10$	0,009	0,368
R^2	0,8303	

Tabel 14 Pendugaan Parameter Model SAR dengan Pengaruh Tetap

Variabel	Koefisien	$p-value$
Rho	0,0138	0,049
$X1$	0,5344	$< 0,000$
$X2$	0,42	0,047
$X3$	0,006	0,626
$X4$	0,007	0,55
$X5$	$-0,065$	0,027
$X6$	0,074	0,048
$X7$	0,041	0,327
$X8$	$-0,012$	0,686
$X9$	$-0,064$	0,026
$X10$	0,008	0,528
R^2	0,8311	

Dari Tabel 12, 13, dan 14, terlihat bahwa tidak ada satu model pun yang semua variabelnya signifikan pada taraf nyata 5%. Tabel 15, 16, dan 17 merupakan hasil dugaan parameter dengan hanya menyertakan variabel-variabel yang signifikan pada setiap modelnya.

Tabel 15 Pendugaan Parameter Model Panel GSM Pengaruh Tetap dengan Variabel-Variabel yang Signifikan

Variabel	Koefisien	$p-value$
Rho	0,253	0,033
$Lambda$	$-0,898$	0,031
$X1$	0,542	$< 0,000$
$X2$	0,409	0,040
$X4$	0,013	0,019
$X5$	$-0,055$	0,028
$X7$	0,083	0,018
$X9$	$-0,055$	0,020
R^2	0,8242	

Tabel 16 Pendugaan Parameter Model Panel SEM Pengaruh Tetap dengan Variabel-Variabel yang Signifikan

Variabel	Koefisien	$p-value$
$Lambda$	$-0,144$	0,028
$X1$	0,549	$< 0,000$
$X2$	0,425	0,039
$X5$	$-0,046$	0,021
$X6$	0,044	0,021
$X7$	0,010	0,017
R^2	0,8204	

Tabel 17 Pendugaan Parameter Model Panel SAR Pengaruh Tetap dengan Variabel-Variabel yang Signifikan

Variabel	Koefisien	$p-value$
Rho	0,066	0,046
$X1$	0,542	$< 0,000$
$X2$	0,405	0,048
$X5$	$-0,047$	0,013
$X6$	0,085	0,017
$X9$	$-0,061$	0,007
R^2	0,8292	

4.5. Evaluasi Model

Berdasarkan hasil di atas, diperoleh bahwa model terbaik dalam penelitian ini adalah model *SAR* dengan pengaruh tetap ($R^2 = 0{,}8292$) dan variabel yang signifikan adalah variabel $X1, X2, X5, X6$, dan $X9$. Pemodelan yang terbentuk pada penelitian ini adalah model SAR dengan pengaruh tetap yaitu:

$$y_{it} = 0{,}066 \sum_{j=1}^{N} w_{ij} y_{it} + 0{,}542\, x_{1it} + 0{,}405\, x_{2it} - 0{,}047\, x_{5it} + 0{,}085\, x_{6it} - 0{,}061\, x_{9it} + \varepsilon_{it}$$

5. Kesimpulan

Model panel spasial terbaik dalam menganalisis faktor-faktor yang memengaruhi kesehatan di Provinsi Papua adalah Penelitian ini menghasilkan model terbaik yaitu model SAR dengan pengaruh tetap. Variabel yang memengaruhi persentase penduduk sakit di Provinsi Papua adalah persentase keluhan kesehatan, rata-rata lama sakit, persentase balita tidak imunisasi BCG, DPT, dan HEPATITIS B. Model yang diperoleh pada penelitian ini memiliki koefisien determinasi sebesar 82,92%.

Daftar Pustaka

[1] Rahayu, T.E. 2010. Pertumbuhan dan Persebaran Penduduk Indonesia (Hasil Sensus Penduduk 2010. https://media.neliti.com/media/publications/49963-ID-pertumbuhan-danpersebaran-penduduk-indonesia.pdf. Tanggal akses 5 November 2019.

[2] Kementerian Kesehatan Republik Indonesia. 2019. Program Indonesia Sehat untuk Capai Tingkat Kesehatan Tertinggi. http://sehatnegeriku.kemkes.go.id/baca/rilis-media/20190521/5530314/program-indonesia-sehat-capai-tingkat-kesehatan-tertinggi/. Tanggal akses 5 Nopember 2019.

[3] Kumparan. 2019. 10 Years Challenge: Melihat Kondisi Kesehatan Masyarakat Indonesia. https://kumparan.com/kumparansains/10-years-challenge-melihat-kondisi-kesehatan-masyarakat-indonesia-1547773566224815409/full. Tanggal akses 5 Nopember 2019.

[4] Kementerian Kesehatan Republik Indonesia. 2018. Hasil Riskesdas 2018. http://www.kesmas.kemkes.go.id/assets/upload/dir_519d41d8cd98f00/files/Hasil-riskesdas-2018_1274.pdf. Tanggal akses 5 Nopember 2019.

[5] Badan Pusat Statistik. 2019. Persentase Rumah Tangga Kumuh Perkotaan Menurut Provinsi. https://www.bps.go.id/dynamictable/2019/10/04/1667/persentase-rumah-tangga-

kumuh-perkotaan-40-ke-bawah-menurut-provinsi-2015-2018.html. Tanggal akses 5 Nopember 2019.

[6] Badan Pusat Statistik. 2019. Persentase Rumah Tangga yang Memiliki Akses terhadap Layanan Sanitasi Layak dan Berkelanjutan Menurut Provinsi. https://www.bps.go.id/dynamictable/2019/10/04/1665/persentase-rumah-tangga-yang-memiliki-akses-terhadap-layanan-sanitasi-layak-dan-berkelanjutan-40-bawah-menurut-provinsi-2015-2018.html. Tanggal akses 5 Nopember 2019.

[7] Badan Pusat Statistik. 2019. Persentase Rumah Tangga yang memiliki Akses terhadap Layanan Sumber Air Minum Layak dan Berkelanjutan Menurut Provinsi. https://www.bps.go.id/dynamictable/2019/10/04/1663/persentase-rumah-tangga-yang-memiliki-akses-terhadap-layanan-sumber-air-minum-layak-dan-berkelanjutan-40-bawah-menurut-provinsi-2015-2018.html. Tanggal akses 5 Nopember 2019.

[8] Badan Pusat Statistik. 2018. Angka Kematian Balita per 1000 Kelahiran Hidup Menurut Provinsi. https://www.bps.go.id/dynamictable/2018/06/06/1457/angka-kematian-balita-per-1000-kelahiran-hidup-menurut-provinsi-2012-dan-2017.html. Tanggal akses 5 Nopember 2019.

[9] Anggraeni, Y. 2012. Analisis Spasial Data Panel untuk Menentukan Faktor-Faktor yang Mempengaruhi Kemiskinan di Provinsi Sumatera Selatan. Skripsi. Institut Pertanian Bogor, Bogor.

[10] Hikmah, Y. 2017. Pemodelan Panel Spasial pada Data Kemiskinan di Provinsi Papua. Statistika. 17 (1): 1-15.

[11] Baltagi, B. 2005. Econometrics Analysis of Panel Data 3rd ed. John Wiley & Sons, England.

[12] Pangestika, S. 2015. Analisis Estimasi Model Regresi Data Panel dengan Pendekatan Common Effect Model (CEM), Fixed Effect Model (FEM), dan Random Effect Model (REM). Skripsi. Universitas Negeri Semarang, Semarang.

[13] Anselin, L. 2009. Spatial Regression. Sage Publications, London.

[14] R. Available CRAN Packages By Name. https://cran.r-project.org/web/packages/available_packages_by_name.html. Tanggal akses 5 Nopember 2019.

Lampiran *Package* dan Fungsi Sintaks R

read.csv	library(sp)
library(plm)	readOGR
plmtest	plot
plm	spplot
summary	coordinates
pFtest	mat2listw
phtest	moran.test
library(rgdal)	library(splm)
library(spdep)	spml
library(raster)	

Journal of Data Analysis: Statistics and Its Applications
Writing Guidelines

1. The page size should be A4.Each page should have clear margins of 4cm (top), 3.5cm (left), 2.5 (right) and 2.5cm (bottom). Pages should not contain page numbers.
2. All articles must contain an abstract in Indonesian and English.
3. The number of pages is maximum 20 pages per article.
4. The title should be followed by a list of all authors' names and their affiliations. The authors' affiliations follow the author list. If there is more than one address then a superscripted number should come at the start of each address. Each author should also have a superscripted number or numbers following their name to indicate which address, or addresses, are the appropriate ones for them. E-mail addresses may be given for any or all of the authors.
5. The abstract follows the list of addresses. The abstract text should be indented 25 mm from the left margin at font size 10 and written without keywords. As the abstract is not part of the text it should be complete in itself; no table numbers, figure numbers, references or displayed mathematical expressions should be included. Abstract is preferably written without any abbreviations and should be suitable for direct inclusion in abstracting services.
6. The text of your article should start on the same page as the abstract and is written in single space, NOT double space. Any Acknowledgments should be placed immediately after the last numbered section of the paper, and any appendices after the Acknowledgments section. Any citations in the text body should be written in a square bracket [1] and following the order [2, 3, 4], etc. In regards to this, the reference section does not necessarily have to be written in alphabetical order.
7. Figures and tables should be numbered serially and positioned (centred on the width of the page) close to where they are mentioned in the text, not grouped together at the end. Each figure and table should have a brief explanatory caption. Figure and table numbers should start from 1, following the order 2, 3, etc, and put on bold. For example, the first figure in section 3 should be written as Figure 1 and NOT Figure 3.1.
8. Equations are located in center with number of the quation which must be called in the text.
9. Authors can submit supplementary data attachments to enhance the online version articles. Supplementary data enhancements typically consist of video clips, animations or supplementary data such as data files, tables of extra information or extra figures.
10. If an acknowledgment is made, it should be included at the very end of the paper before the references. This section includes acknowledgment of people, grant details, funds, etc.
11. the authors, in the form: family name (only the first letter capitalized) followed by initials with no periods after the initials; the year of publication; the article title (optional) in lower case letters, except for an initial capital; the journal title (italic and abbreviated). Parts denoted by letters should be inserted after the journal in Roman type; the volume number in bold type; the article number or the page numbers.

    Primer Artikel Primer

    Format: <authors>. <year>. <paper title >. <journal title/periodical> <volume> <(issue)>: <page>.

    For example:

    [1] Harder, S. 2010. From Limestone to Catalysis: Application of Calcium Compounds as Homogeneous Catalysts. Chem. Rev. 110 (7): 3852– 3876.

    Text Book

    Format: <authors>. <year>. <*book title*>. <pubisher>, <city>.

    For Example:

    [2] Stanbury, P. F., Whitaker, A., and Hall, S. J. 1995. Principles of Fermentation Technology 2nd ed. Butterworth-Heinemann, Oxford UK.

ISSN 2623-2286 (On-line); ISSN 2623-0658 (Print)