



## Analisis Kluster untuk Pengelompokan Kecamatan di Kabupaten Karo Berdasarkan Indikator Pendidikan

Asbina Ginting<sup>1</sup>, Marline S. Paendong<sup>\*</sup>, Djoni Hatidja<sup>1</sup>

<sup>1</sup>Jurusan Matematika–Fakultas Matematika dan Ilmu Pengetahuan Alam–Universitas Sam Ratulangi Manado, Indonesia

<sup>\*</sup>Corresponding author : [marlinepaendong@unsrat.ac.id](mailto:marlinepaendong@unsrat.ac.id)

### ABSTRAK

Analisis kluster merupakan salah satu metode pengelompokan objek yang memiliki kedekatan/kemiripan karakteristik. Metode yang digunakan pada analisis kluster yaitu metode hirarki yaitu single linkage dengan ukuran jarak euclidean. Tujuan dari penelitian ini untuk mengelompokkan kecamatan di Kabupaten Karo berdasarkan Indikator Pendidikan. Data yang digunakan adalah data sekunder yang diperoleh dari website Badan Pusat Statistik (BPS) Kabupaten Karo. Hasil penelitian pengelompokan kecamatan di Kabupaten Karo sebanyak 3 kluster. Kluster pertama terdiri dari 15 kecamatan yaitu Kutabuluh, Juhar, Naman Teran, Tiganderket, Simpang Empat, Merdeka, Dolat Rayat, Payung, Merek, Mardinding, Laubaleng, Barusjahe, Munte, Tigapanah, dan Tiga Binanga dengan tingkat pendidikan yang rendah. Kluster kedua terdiri dari 1 kecamatan yaitu Kecamatan Berastagi dengan tingkat pendidikan sedang. Kluster ketiga terdiri dari 1 kecamatan yaitu Kecamatan Kabanjahe dengan tingkat pendidikan tinggi.

### INFO ARTIKEL

Diterima : 0000

Diterima setelah revisi : 0000

Tersedia online : 0000

### Kata Kunci:

Pengelompokan  
Indikator Pendidikan  
Jarak Euclidean  
Single Linkage

### ABSTRACT

Cluster analysis is a method of grouping objects that have closeness/similar characteristics. The method used in cluster analysis is the hierarchical method, namely single linkage with a Euclidean distance measure. The aim of this research is to group sub-districts in Karo Regency based on Education Indicators. The data used is secondary data obtained from the Karo Regency Central Statistics Agency (BPS) website. The results of the research into 3 clusters of sub-districts in Karo Regency. The first cluster consists of 15 sub-districts, namely Kutabuluh, Juhar, Naman Teran, Tiganderket, Simpang Empat, Merdeka, Dolat Rayat, Payung, Brand, Mardinding, Laubaleng, Barusjahe, Munte, Tigapanah, and Tiga Binanga with low levels of education. The second cluster consists of 1 sub-district, namely Berastagi District with a medium level of education. The third cluster consists of 1 sub-district, namely Kabanjahe District with a high level of education.

### ARTICLE INFO

Accepted : 0000

Accepted after revision : 0000

Available online : 0000

### Keywords:

Grouping  
Education Indicators  
Euclidean Distance  
Single Linkage

### 1. PENDAHULUAN

Salah satu parameter keberhasilan pendidikan dapat dilihat dari indikator pendidikan di suatu daerah. Pentingnya pendidikan telah mendorong pemerintah dan lembaga terkait untuk mengumpulkan data indikator pendidikan di seluruh wilayah, termasuk di tingkat kecamatan. Data tersebut mencakup berbagai variabel seperti tingkat akses pendidikan, kualitas pengajaran dan tingkat partisipasi siswa. Namun, dalam mengelola wilayah yang memiliki banyak kecamatan, diperlukan strategi yang bijak untuk mengidentifikasi pola dan pendidikan antara kecamatan-kecamatan tersebut.

Kabupaten Karo merupakan salah satu kabupaten yang ada di Sumatera Utara Indonesia, yang terdiri dari

17 Kecamatan, yang berkomitmen untuk mengembangkan pendidikan sebagai upaya untuk meningkatkan sumber daya manusia ke arah yang lebih baik terhadap daerah ini. Namun, meskipun memiliki komitmen yang tinggi, tingkat pendidikan di Kabupaten Karo mungkin menghadapi tantangan dan karakteristik yang berbeda dalam sektor pendidikan.

Untuk tujuan analisis pengelompokan kecamatan berdasarkan indikator-indikator pendidikan, dapat digunakan metode Analisis Peubah Ganda (APG) atau Multivariate Analysis. Di dalam analisis peubah ganda, salah satu analisis yang dikenal dengan kemampuan mengelompokkan adalah Analisis kluster (Cluster Analysis).

Analisis kluster adalah salah satu dari metode dalam analisis multivariat yang memiliki tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya (Sari & Sukestiyarno, 2021). Analisis kluster dapat membantu dalam mengelompokkan kecamatan-kecamatan berdasarkan indikator pendidikan tersebut. Pengelompokan ini dapat memberikan pemahaman yang lebih mendalam tentang distribusi pendidikan di wilayah tersebut, membantu pemerintah dan pemangku kepentingan untuk mengidentifikasi kebijakan pendidikan yang perlu ditingkatkan

### Pendidikan

Pendidikan merupakan “usaha sadar dan terencana untuk mewujudkan suasana belajar dan pembelajaran yang menyenangkan agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya dan masyarakat”. Pendidikan meliputi pengajaran keahlian khusus, dan juga sesuatu yang tidak dapat dilihat tetapi lebih mendalam yaitu pemberian pengetahuan, pertimbangan dan kebijaksanaan (Pristiwanti, et al, 2022).

Pendidikan di Indonesia mengacu pada delapan standar pendidikan yang dinamakan Standar Nasional Pendidikan (SNP) yaitu terdiri dari standar kompetensi lulusan, standar isi, standar proses, standar pengelolaan, standar pendidik dan tenaga kependidikan, standar evaluasi, standar pembiayaan, standar sarana dan prasarana. SNP dikembangkan dan ditetapkan untuk mengukur, mengevaluasi, menilai mutu pendidikan, dalam rangka meningkatkan mutu pendidikan. Masing masing komponen dalam SNP saling terkait dan membentuk sebuah sistem penyelenggaraan pendidikan mulai dari input, proses serta output (Alawiyah, 2017).

### Analisis Deskriptif

Analisis deskriptif adalah teknik statistik yang digunakan untuk menganalisa data dengan cara mendeskripsikan, menyederhanakan, dan menyajikan data sampel ke dalam bentuk yang teratur supaya mudah dipahami (Gulla, et al, 2023). Pada analisis deskriptif data biasanya ditampilkan dalam bentuk tabel biasa atau tabel frekuensi, grafik, diagram batang, diagram garis, diagram lingkaran, ukuran pemusatan data, ukuran penyebaran data dan sebagainya (Sugiyono, 2010).

### Analisis Kluster

Analisis kluster merupakan suatu metode yang membagi individu ke dalam kelompok yang bermakna. Analisis kluster mengelompokkan objek berdasarkan informasi yang diperoleh pada data yang menggambarkan objek dan keterkaitannya (Johnson dan Wichern, 2002).

Analisis kluster terdiri dari dua metode yaitu metode hirarki (hierarchical clustering method), dan metode non hirarki (nonhierarchical clustering method). Metode hirarki digunakan apabila banyaknya gerombol yang akan dibentuk belum diketahui sebelumnya, sedangkan metode non hirarki bertujuan untuk mengelompokkan objek ke dalam k gerombol tertentu ( $k$

<  $n$ ), di mana nilai  $k$  telah ditentukan sebelumnya (Pamekas, et al, 2019).

Metode analisis kluster membutuhkan suatu ukuran ketakmiripan (jarak) yang didefinisikan untuk setiap pasang objek yang akan dikelompokkan. Pada penelitian ini jarak yang akan digunakan adalah jarak Euclidean. Jarak Euclidean adalah besarnya jarak suatu garis lurus yang menghubungkan antar objek. Jarak Euclidean merupakan akar jumlah kuadrat perbedaan nilai untuk tiap variabel. Jarak Euclidean dirumuskan sebagai berikut (Nugroho, 2008).

$$d_{i,j} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (3)$$

Keterangan :

$d_{i,j}$  = jarak antar objek  $i$  dengan objek  $j$

$x_{ik}$  = nilai objek  $i$  pada variabel ke- $k$

$x_{jk}$  = nilai objek  $j$  pada variabel ke- $k$

$p$  = banyaknya variabel diamati

### Metode Hirarki

Ada empat metode aglomeratif dalam pembentukan kluster menurut Simamora (2005) yaitu : metode single linkage, metode complete linkage, metode average linkage, dan metode ward. Metode yang digunakan dalam penelitian ini adalah metode single linkage.

### Single linkage

Metode pautan tunggal adalah proses pengklasteran yang didasarkan pada jarak terdekat antar objek. Jika dua objek terpisah oleh jarak yang pendek, maka kedua objek tersebut akan digabung menjadi satu kluster. Metode ini dimulai dengan menentukan objek-objek yang memiliki jarak terkecil dalam matriks *proximity*  $D = \{d_{ik}\}$ . Objek-objek yang memiliki jarak terkecil digabungkan dalam satu kluster, misal  $U$  dan  $V$  adalah dua objek yang akan diklasterkan sehingga diperoleh kluster  $(UV)$ . Untuk mencari jarak antara kluster  $(UV)$  dan kluster  $W$  atau kluster lainnya diperoleh dengan rumus:

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (2)$$

Keterangan :

$d_{UW}$  adalah jarak terdekat dari gerombol  $U$  dan  $W$

$d_{VW}$  adalah jarak terdekat dari gerombol  $V$  dan  $W$

## 2. METODE PENELITIAN

### Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari website Badan Pusat Statistik (BPS) Kabupaten Karo, yaitu data Kabupaten Karo dalam angka tahun 2023 tentang pendidikan.

### Metode Analisis

Adapun langkah-langkah analisis data dalam pengelompokan kecamatan di kabupaten karo berdasarkan indikator Pendidikan dalah sebagai berikut:

1. Pengambilan data sekunder
2. Melakukan analisis deskriptif variable-variabel penelitian
3. Melakukan pengolahan data menggunakan analisis kluster dengan metode single linkage

4. Menentukan banyaknya klaster
5. Membuat kesimpulan

### Variabel Penelitian

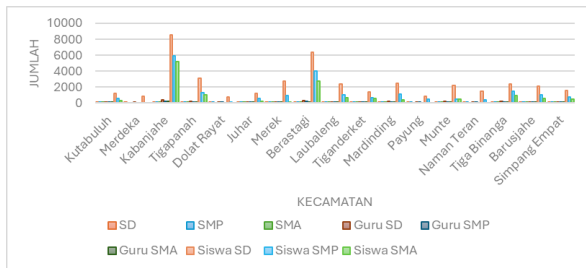
**Tabel 1. Variabel Penelitian**

No.	Variabel	Keterangan
1	$X_1$	Banyak Sekolah Dasar (SD)
2	$X_2$	Banyak Sekolah Menengah Pertama (SMP)
3	$X_3$	Banyak Sekolah Menengah Atas (SMA)
4	$X_4$	Banyak Guru SD
5	$X_5$	Banyak Guru SMP
6	$X_6$	Banyak Guru SMA
7	$X_7$	Banyak Siswa SD
8	$X_8$	Banyak Siswa SMP
9	$X_9$	Banyak Siswa SMA

### 3. HASIL DAN PEMBAHASAN

#### Analisis Deskriptif

Berdasarkan data indikator Pendidikan di setiap kecamatan analisis deskriptif dapat dilihat pada diagram yang disajikan pada gambar 1:



**Gambar 1.** Data indikator Pendidikan di setiap Kecamatan

Berdasarkan gambar 11 dapat dilihat bahwa jumlah siswa SD paling banyak ada di 2 Kecamatan, yaitu Kecamatan kabanjahe dan Kecamatan Berastagi. Kecamatan Merdeka dan kecamatan Dolat Rayat memiliki jumlah indikator pendidikan yang paling rendah dibandingkan kecamatan lainnya.

#### Pengukuran Jarak

Dalam menghitung kemiripan tiap objek (kecamatan), metode yang digunakan adalah jarak euclidean. Terdapat 17 kecamatan yang akan dihitung kemiripannya.

Misalnya menghitung kemiripan antara Kutabuluh dengan kecamatan Merdeka (objek 1 dan 2)

$$d_{i,j} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

$$d_{(1,2)} = \sqrt{\frac{((14-6)^2 + (3-0)^2 + (1-0)^2 + (108-56)^2 + (61-0)^2 + (25-0)^2 + (1257-875)^2 + (586-0)^2 + (302-0)^2)}{343396 + 91204}}$$

$$= \sqrt{\frac{64 + 9 + 1 + 1704 + 3721 + 625 + 145924 + 343396 + 91204}{71289 + 12996}}$$

$$= 766,582$$

Kemudian mengitung kemiripan antara kecamatan kutabuluh dengan kecamatan kabanjahe (objek 1 dan objek 3).

$$d_{(1,3)} = \sqrt{\frac{((14-36)^2 + (3-16)^2 + (1-9)^2 + (108-426)^2 + (61-254)^2 + (25-254)^2 + (1257-8530)^2 + (586-5891)^2 + (302-5229)^2)}{28143025 + 24275329}}$$

$$= \sqrt{\frac{484 + 169 + 64 + 101124 + 37249 + 52441 + 52896529 + 28143025 + 24275329}{71289 + 12996}}$$

$$= 10271,63$$

Dan seterusnya sampai menghitung kemiripan antara kecamatan Simpang Empat dengan kecamatan Barusjahe (Objek 17 dan 16).

$$d_{(17,16)} = \sqrt{\frac{((14-21)^2 + (2-5)^2 + (1-1)^2 + (121-177)^2 + (63-91)^2 + (26-34)^2 + (1579-2127)^2 + (737-1004)^2 + (453-567)^2)}{71289 + 12996}}$$

$$= \sqrt{\frac{49 + 9 + 0 + 3136 + 784 + 64 + 300304 + 71289 + 12996}{71289 + 12996}}$$

$$= 623,40$$

Dari perhitungan tersebut dapat dilihat hasil perhitungan jarak euclidean antara Kecamatan Kutabuluh dengan Kecamatan Merdeka adalah 766,582. sedangkan jarak euclidean antara Kecamatan Kutabuluh dengan Kecamatan Kabanjahe adalah 10271,63. Hal ini menunjukkan bahwa Kecamatan Kutabuluh memiliki karakteristik yang lebih mirip dengan kecamatan Merdeka daripada Kecamatan Kutabuluh dengan kecamatan Kabanjahe. Demikian dengan penafsiran objek yang lainnya, semakin kecil jarak antara kedua objek maka akan semakin mirip karakteristik dari kedua objek tersebut (Is, et al, 2021).

#### Hasil Analisis Klaster

##### Melakukan Perbaikan Matriks Jarak

Dibawah ini merupakan proses pengelompokan secara manual dengan metode *single linkage*. Setelah melakukan perhitungan jarak yang dijadikan sebagai ukuran kedekatan atau kemiripan pengelompokan dimulai dari jarak terdekat, maka perhitungan dimulai dari objek 1 (kecamatan Kutabuluh) dan objek 6 (Kecamatan Juhar), dengan jarak sebesar 99,40.

$$d_{(1,6)1} = \{d_{(1)1}, d_{(6)1} = \{0, 99,80\} = 0$$

$$d_{(1,6)2} = \{d_{(1)2}, d_{(6)2} = \{766.58, 728.40\} = 728.40$$

⋮

⋮

$$d_{(1,6)17} = \{d_{(1)17}, d_{(6)17} = \{386.60, 447.16\} = 386.60$$

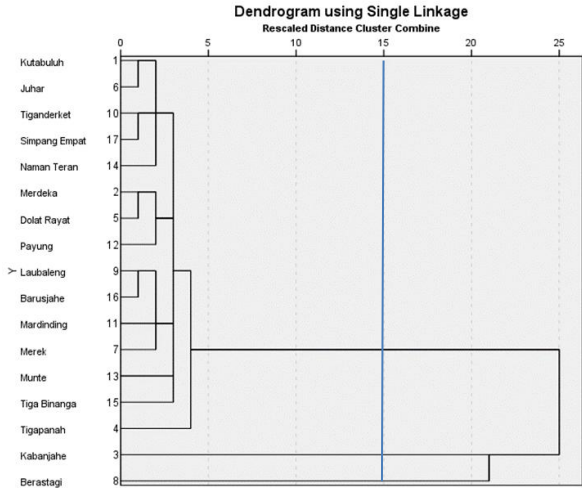
Perhitungan kecamatan Kutabuluh akan dilakukan untuk semua kecamatan kecuali kecamatan Kutabuluh dan kecamatan Juhar itu sendiri. Jarak terkecil adalah jarak yang akan diambil untuk disatukan menjadi satu kelompok. Seterusnya perhitungan dilakukan sampai matriks jarak dari semua objek yang telah digabungkan menghasilkan matriks baru. Sehingga untuk hasil keseluruhan matriks baru adalah sebagai berikut:

**Tabel 2.** Matriks baru metode zingle linkage

Kelompok			
	1,6,10,17,14,2,5,12,9,16,11,7,13,15,4	3	8
1,6,10,17,14,2,5,12,9,11,16,7,13,15,4	0	8240,67	4567,20
3	8240,67	0	3773,54
8	4567,20	3773,54	0

### Menentukan jumlah kluster dan anggotanya

Proses pengelompokan juga dapat diilustrasikan dalam bentuk dendrogram. Dendrogram pada metode single linkage dapat dilihat pada Gambar 2.



**Gambar 2.** Dendrogram metode single linkage

Gambar 12 menunjukkan hasil dendrogram yang didapatkan dari perpotongan pada jarak 15 yaitu sebanyak 3 kelompok. Perpotongan dilakukan untuk mengetahui kecamatan apa saja yang masuk pada masing-masing pengelompokan. Hasil perpotongan tersebut disajikan pada tabel 3.

**Tabel 3.** Anggota dari Klaster yang Terbentuk

Klaster	Anggota
I	Kutabuluh, Juhar, Naman Teran, Tiganderket, Simpang Empat, Merdeka, Dolat Rayat, Payung, Merek, Mardinding, Laubaleng, Barusjahe, Munte, Tigapanah, Tiga Binanga
II	Berastagi
III	Kabanjahe

### Karakteristik Kelompok

**Tabel 4.** Karakteristik Kelompok

Variabel	Rata-rata		
	Kelompok 1	Kelompok 2	Kelompok 3
X1	15,73	26	36
X2	2,87	9	16
X3	0,80	4	9
X4	135,47	326	426
X5	55,87	236	254
X6	24,47	141	254
X7	1788,33	6371	8530
X8	733,67	3983	5891
X9	383,67	2797	5229

Berdasarkan nilai rata-rata variabel untuk setiap kelompok maka dapat diketahui karakteristik dari 3

kelompok yang terbentuk, sehingga diperoleh sebagai berikut:

1. Kelompok 1 pada variabel X3 menunjukkan bahwa jumlah sekolah menengah atas (SMA) memiliki nilai rata-rata yang paling rendah dibandingkan dengan variabel lainnya. Dalam kelompok ini, terdapat jumlah sekolah SMA yang paling minim dibandingkan dengan kelompok lainnya. Kelompok 1 ini terdiri dari Kutabuluh, Juhar, Naman Teran, Tiganderket, Simpang Empat, Merdeka, Dolat Rayat, Payung, Merek, Mardinding, Laubaleng, Barusjahe, Munte, Tigapanah, dan Tiga Binanga, yang memiliki tingkat pendidikan yang rendah. Kecamatan Merdeka dan Dolat Rayat menonjol karena memiliki jumlah indikator pendidikan yang paling rendah dibandingkan dengan kecamatan lainnya. Ke-15 kecamatan tersebut dikelompokkan menjadi Kelompok 1 karena memiliki jumlah sekolah dasar (SD), sekolah menengah pertama (SMP), dan SMA yang terendah. Kehadiran jumlah sekolah yang minim di wilayah ini juga berdampak pada jumlah guru dan siswa yang rendah.
2. Kelompok 2 terdiri dari Kecamatan Berastagi dengan tingkat pendidikan yang sedang, di mana memiliki jumlah sekolah, guru, dan siswa yang lebih besar dibandingkan dengan kelompok 1.
3. Kelompok 3 memperoleh nilai tertinggi di antara kedua kelompok lainnya. Hal ini dapat terbukti dari variabel X7, X8, dan X9 yang menunjukkan nilai rata-rata yang tinggi. Kelompok 3 terdiri dari Kecamatan Kabanjahe dengan tingkat pendidikan yang tinggi, serta memiliki jumlah sekolah, guru, dan siswa yang tinggi.

### 4. PENUTUP Kesimpulan

Berdasarkan hasil penelitian yang dilakukan pada 17 kecamatan di Kabupaten Karo dikelompokkan dalam 3 klaster. Klaster 1 terdiri dari 15 kecamatan dengan tingkat pendidikan yang rendah, termasuk Kutabuluh, Juhar, Naman Teran, Tiganderket, Simpang Empat, Merdeka, Dolat Rayat, Payung, Merek, Mardinding, Laubaleng, Barusjahe, Munte, Tigapanah, dan Tiga Binanga. Klaster 2 hanya terdiri dari 1 kecamatan yaitu Berastagi, dengan tingkat pendidikan yang sedang. Sementara itu, klaster 3 juga hanya terdiri dari 1 kecamatan, yaitu Kabanjahe, dengan tingkat pendidikan yang tinggi. Kecamatan Kabanjahe memiliki jumlah nilai tertinggi untuk semua variabel. Jumlah siswa SD (variabel X7) mencapai nilai tertinggi, sedangkan jumlah Sekolah Menengah Atas (SMA) (variabel X3) memiliki nilai terendah.

### REFERENSI

- [1] Alawiyah, F. (2017). Standar Nasional Pendidikan Dasar dan Menengah. *Aspirasi*, 8(1), 81-92.
- [2] Badan Pusat Statistik (BPS) Kabupaten Karo Dalam Angka 2023.  
<https://karokab.bps.go.id/publikasi.html> [diakses pada 16 Oktober 2023]
- [3] Gulla, A. M., Komalig, H. A., & Salaki, D. T. (2023). Analisis Gerombol untuk Pengelompokan Program Studi di Fmipa Unsrat Berdasarkan Keadaan Orang Tua. *d'CartesiaN: Jurnal Matematika dan Aplikasi*, 12(2), 63-66.
- [4] Johnson, R. A., and D. W. Wichern. 2002. *Applied Multivariate Statistical Analysis*, Fifth Edition. Pearson Education. United States Of America
- [5] Ls, D., Talakua, M. W., Lesnussa, Y. A., & Matdoan, M. Y. (2021). Analisis Klaster untuk Pengelompokan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Pendidikan dengan Menggunakan Metode ward. *Statistika dan Aplikasinya*, 5(1), 51-60.
- [6] Nugroho, S. 2008. *Statistika Multivariat terapan*, edisi pertama. UNIB Press, Universitas Bengkulu.
- [7] Pamekas, J. A., Prang, J. D., & Hatidja, D. (2019). Pengelompokan dosen berdasarkan proses pembelajaran di program studi matematika fmipa unsrat pada semester genap 2017/2018 menggunakan metode ward. *d'CartesiaN : Jurnal Matematika dan Aplikasi*, 19(1), 47-56.
- [8] Pristiwanti, D., Badariah, B., Hidayat, S., & Dewi, R. S. (2022). Pengertian Pendidikan. *Pendidikan dan Konseling*, 4(6), 7911-7915.
- [9] Sari, D. N., & Sukestiyarno, Y. (2021). Analisis Cluster dengan Metode K-Means pada Persebaran. *PRIMSA*, 603-610.
- [10] Simamora, Bilson. (2005). *Analisis Multivariat Pemasaran Edisi Pertama*. Jakarta: PT. Gramedia Pustaka tama.



**Asbina Br Ginting** ([asbinaginting3@gmail.com](mailto:asbinaginting3@gmail.com))

Lahir di Bulanjulu, Sumatera Utara pada tanggal 05 September 2001. Menempuh pendidikan tinggi Jurusan Matematika, FMIPA, Universitas Sam Ratulangi Manado. Tahun 2024 adalah tahun terakhir ia menempuh studi. Makalah ini merupakan hasil penelitian skripsinya yang dipublikasikan.

**Marline S. Paendong** ([marlinepaendong@unsrat.ac.id](mailto:marlinepaendong@unsrat.ac.id))



Pada tahun 1999, memperoleh gelar sarjana di Program Studi Matematika, Universitas Gadjah Mada dengan gelar Sarjana Sains (S.Si.). Pada tahun 2006, gelar Magister Sains diperoleh di Institut Pertanian Bogor. Ia bekerja di Program Studi Matematika, UNSRAT sebagai dosen dengan jabatan fungsional sebagai Lektor.

**Djoni hatidja** ([dhatidja@unsrat.ac.id](mailto:dhatidja@unsrat.ac.id))



Lahir di Minahasa Tenggara, 16 Juli 1969. Gelar sarjana diperoleh tahun 1994 di Program Studi Statistika IPB, Bogor. Tahun 2000 menyelesaikan studi s2 di program Studi Statistika Program Pasca Sarjana IPB, Bogor. Saat ini mengajar di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sam Ratulangi Manado.