

# JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 13, NOMOR 2, DESEMBER 2021 ISSN 2086 – 4132  
AKREDITASI KEMENRISTEK-DIKTI NOMOR: 9 TAHUN 2018

Penerapan *Social Network Analysis* dan *Latent Dirichlet Allocation* Pada Publikasi Penelitian Dosen

MUHAMMAD IQBAL dan SETIA PRAMANA

Interpolasi Curah Hujan Ekstrem Menggunakan Model *Spatial* Di Provinsi Jawa Barat

RAHMI LATHIFAH ISLAMI dan PARDOMUAN ROBINSON SIHOMBING

Analisis Jumlah Penduduk Miskin di Provinsi Maluku dengan Menggunakan Pendekatan Regresi Spasial

SALMON NOTJE AULELE, V. Y. I. ILWARU, E. R. WURITIMUR, dan M. Y. MATDOAN

Pembentukan Portofolio Saham Menggunakan Klastering *Time Series* K-Medoid dengan Ukuran Jarak *Dynamic Time Warping*

LA GUBU, DEDI ROSADI dan ABDURAKHMAN ABDURAKHMAN

*Sex and Age Group Differences in The Spread of Tuberculosis In Indonesia: An Agent Based Modeling Approach*

DWI ARI SURYAWAN. S dan TIODORA HADUMAON SIAGIAN

*Robust Biplot Analysis of Natural Disasters in Indonesia from 2019 to 2021*

HILDA VENELIA, KHOIRIN NISA, RIZKI AGUNG WIBOWO, dan MONA ARIF MUDA



PUSAT PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT  
POLITEKNIK STATISTIKA STIS

# JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

---

Jurnal “Aplikasi Statistika dan Komputasi Statistik” memuat karya ilmiah hasil penelitian dan kajian teori statistik dan komputasi statistik yang diterapkan khususnya pada bidang ekonomi dan sosial kependudukan, serta teknologi informasi yang terbit dua kali dalam setahun setiap bulan Juni dan Desember.

<b>Penanggung Jawab:</b>	Dr. Erni Tri Astuti
<b>Ketua Dewan Redaksi:</b>	Dr. Nasrudin
<b>Koordinator Jurnal Ilmiah:</b>	Dr. Ernawati Pasaribu
<b>Mitra Bestari:</b>	Dr. Azka Ubaidillah Dr. Cucu Suamarni Dr. Margaretha Ari Anggorowati Dr. Rita Yuliana Dr. Timbang Sirait Dr. Waris Marsisno
<b>Pelaksana Redaksi:</b>	Lutfi Rahmatuti Maghfiroh, SST., M.T. Geri Yesa Ermawan, S.Tr.Stat. Rahadi Jalu Yoga Utama, S.Tr.Stat.

## **Alamat Redaksi:**

Politeknik Statistika STIS  
Jl. Otto Iskandardinata 64C  
Jakarta Timur 13330  
Telp. 021-8191437

Redaksi menerima karya ilmiah atau artikel penelitian mengenai kajian teori statistik dan komputasi statistik pada bidang ekonomi dan sosial kependudukan, serta teknologi informasi. Redaksi berhak menyunting tulisan tanpa mengubah makna substansi tulisan. Isi Jurnal Aplikasi Statistika dan Komputasi Statistik dapat dikutip dengan menyebutkan sumbernya.

## **PENGANTAR REDAKSI**

Puji syukur kehadiran Allah, Tuhan Yang Maha Esa, “Jurnal Aplikasi Statistika dan Komputasi Statistik” Volume 13, Nomor 1, Juni 2021 dapat diterbitkan. Jurnal ilmiah ini dapat terwujud atas partisipasi semua pihak, penulis internal dilingkungan Politeknik Statistika STIS maupun penulis eksternal, serta mitra bestari.

Semoga artikel dalam jurnal ini dapat menambah pengetahuan para pembaca tentang penggunaan metode statistika serta komputasi statistik pada berbagai jenis data. Redaksi terus menunggu artikel-artikel ilmiah selanjutnya dari Bapak/Ibu agar publikasi yang dihasilkan menjadi salah satu sarana untuk memberikan sosialisasi statistika bagi masyarakat.

Jakarta, Juni 2021  
Ketua Dewan Redaksi,

**Dr. Nasrudin**

# JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 13, NOMOR 1, JUNI 2021

AKREDITASI KEMENRISTEK-DIKTI NOMOR: 9 TAHUN 2018

## DAFTAR ISI

<b>Pengantar Redaksi</b> .....	iii
<b>Daftar Isi</b> .....	iv
<b>Abstrak</b> .....	v-xii
<b>Kajian Usaha Rumah Tangga Budidaya Bandeng di Sulawesi Selatan Tahun 2013</b> <u>Muhamad Saiful Hadi dan Putri Lydia Eltheofany S.</u> .....	1-8
<b>Pemodelan Sebaran <i>Total Dissolved Solid</i> Menggunakan Metode <i>Mixed Geographically Weighted Regression</i></b> <u>Dadan Kusnandar, Kusnandar, Naomi Nessyana Debataraja, dan Siti Fitriani</u> .....	9-16
<b>Penerapan Peta Kendali Atribut Klasik dan Peta Kendali NP Bayes pada Produk Cacat Air Minum Asri di CV. Multi Rejeki Selaras Payakumbuh</b> <u>Ferra Yanuar, Mutiara Fara Nabilla, dan Izzati Rahmi</u> .....	17-24
<b><i>Small Area Estimation with Excess Zero</i> (Studi Kasus: Angka Kematian Bayi di Pulau Jawa)</b> <u>Nofita Istiana</u> .....	25-34
<b>Pemodelan Kerugian Bencana Banjir Akibat Curah Hujan Ekstrem Menggunakan EVT dan <i>Copula</i></b> <u>Ian Surya Prayoga dan Atina Ahdika</u> .....	35-46
<b>Pengelompokkan Provinsi di Indonesia Berdasarkan Jumlah Kasus Covid-19 dan Fasilitas Kesehatan</b> <u>Meinisa Fadillah Rahmi, Paulus Satria Prasetyo, Ratih Nurhabibah, Rizky Perdana, dan Wa Ode Zuhayeni Madjida</u> .....	47-56

Kata kunci bersumber dari artikel. Lembar abstrak ini boleh diperbanyak tanpa izin dan biaya

DDC: 315.98

Muhamad Saiful Hadi dan Putri Lydia  
Eltheofany S.

Kajian Usaha Rumah Tangga Budidaya  
Bandeng di Sulawesi Selatan Tahun 2013

Jurnal Aplikasi Statistika & Komputasi  
Statistik, Volume 13, Nomor 1, Juni 2021,  
hal 1 – 8

Kata kunci: budidaya bandeng, regresi  
linier berganda, *return to scale*, skala  
ekonomis

#### Abstrak

Sejak tahun 2010-2013 produksi bandeng di Indonesia merupakan tertinggi di dunia. Produksi bandeng tertinggi di Indonesia yaitu di Sulawesi Selatan. Namun, pertumbuhan produksi bandeng di provinsi tersebut dari tahun 2010-2012 mengalami penurunan dari 20,67% menjadi 2,75%. Hal tersebut mengindikasikan bahwa pelaku usaha budidaya bandeng masih kurang memaksimalkan variabel input untuk produksi bandeng. Penelitian ini bertujuan untuk menganalisis pengaruh luas lahan, biaya benih, biaya pupuk dan obat-obatan, dan biaya tenaga kerja terhadap produksi bandeng, menganalisis *Return to Scale*, dan menganalisis skala ekonomis di provinsi Sulawesi Selatan tahun 2013. Penelitian menggunakan metode analisis regresi linier berganda. Berdasarkan hasil analisis diketahui bahwa luas lahan, biaya benih, biaya pupuk dan obat-obatan, dan biaya tenaga kerja berpengaruh signifikan dan positif dengan tingkat kepercayaan 95% terhadap produksi bandeng di Sulawesi Selatan tahun 2013. Usaha budidaya bandeng di Sulawesi Selatan tahun 2013 berada pada kondisi *Increasing Return to Scale*. Persentase rumah tangga yang berada pada kondisi skala ekonomis sebesar 69,53%.

DDC: 315.98

Dadan Kusnandar, Kusnandar, Naomi Nesyana Debataraaja, dan Siti Fitriani

Pemodelan Sebaran *Total Dissolved Solid* Menggunakan Metode *Mixed Geographically Weighted Regression*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 13, Nomor 1, Juni 2021, hal 9 – 16

Abstrak

Model mixed geographically weighted regression (Mixed GWR) merupakan teknik mengeksplorasi ketidakstasioneran spasial dengan mempertimbangkan pengaruh lokal dan global secara bersamaan. Dalam penelitian ini, Mixed GWR digunakan untuk menentukan model sebaran Total Dissolved Solid (TDS) pada kualitas air di Kota Pontianak. Variabel independen yang digunakan yaitu Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD), pH, kesadahan dan warna. Hasil penelitian menunjukkan model Mixed GWR dengan pembobot Fixed Bivariate Kernel lebih baik dibandingkan model regresi global dan GWR dengan AIC sebesar 326,48 dan MAPE 22,34%, dan variabel yang berpengaruh secara global yaitu kesadahan dan warna, sedangkan lokal yaitu COD, BOD dan pH.

Kata kunci: pencemaran air, regresi spasial, regresi global

DDC: 315.98

Ferra Yanuar, Mutiara Fara Nabilla, dan Izzati Rahmi

Penerapan Peta Kendali Atribut Klasik dan Peta Kendali NP Bayes pada Produk Cacat Air Minum Asri di CV. Multi Rejeki Selaras Payakumbuh

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 13, Nomor 1, Juni 2021, hal 17 – 24

Abstrak

Usaha yang dapat dilakukan untuk menekan jumlah produk yang cacat atau rusak adalah dengan melakukan pengendalian kualitas. Pengendalian kualitas yang baik akan membantu dalam kelancaran proses produksi. Salah satu alat statistik yang dapat digunakan untuk mengevaluasi apakah suatu proses produksi berada dalam keadaan terkendali atau tidak secara statistik adalah peta kendali. Pada penelitian ini pengidentifikasian kualitas dilakukan dengan membuat peta kendali atribut yaitu peta kendali np. Peta kendali np ini dikonstruksi berdasarkan distribusi Binomial yang kemudian akan diduga menggunakan metode Bayes sehingga diperoleh peta kendali np Bayes. Selanjutnya, untuk menguji kinerja dari peta kendali ini, dilakukan perbandingan kedua bagan kendali berdasarkan nilai Average Run Length (ARL). Hasil analisis dan pembahasan diperoleh bahwa peta kendali np Bayes prior konjugat lebih sensitif dalam mendeteksi data diluar kendali daripada peta kendali np Bayes prior non informatif. Peta kendali ini juga memiliki kinerja yang lebih baik karena memiliki nilai ARL yang lebih kecil dibandingkan peta kendali np klasik.

Kata kunci: Peta kendali np, metode Bayes, *Average Run Length* (ARL)

DDC: 315.98

Nofita Istiana

*Small Area Estimation with Excess Zero*  
(Studi Kasus: Angka Kematian Bayi di Pulau Jawa)

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 13, Nomor 1, Juni 2021, hal 25 – 34

Abstrak

Angka Kematian Bayi (AKB) merupakan indikator yang penting karena dapat digunakan untuk membandingkan status kesehatan antar penduduk. Pendugaan AKB berdasarkan Survei Demografi dan Kesehatan Indonesia (SDKI) hanya berskala nasional atau provinsi namun dengan adanya sistem desentralisasi diperlukan prediksi untuk area yang lebih kecil seperti kabupaten/kota. *Small Area Estimation* (SAE) dapat digunakan untuk menduga AKB di tingkat kabupaten/kota, yaitu dengan menggunakan *generalized linear mixed* model. AKB merupakan data cacahan dengan peluang sangat kecil, sehingga diasumsikan mengikuti sebaran Poisson. Model Poisson memiliki asumsi rata-rata sama dengan ragam, tetapi asumsi ini sering terlanggar. Salah satu penyebabnya yaitu proporsi nilai nol yang sangat besar pada data (*excess zero*). Penelitian ini menggunakan *Zero Inflated Poisson (ZIP) mixed model* sebagai alternatif dari *Poisson mixed model*. Hasil perbandingan RMSE, *ZIP mixed model* jauh lebih baik dibandingkan *Poisson mixed model* serta dapat memperbaiki hasil dugaan langsung AKB.

Kata kunci: SAE, AKB, *excess zero*, *count data*, *ZIP mixed model*

DDC: 315.98

Ian Surya Prayoga dan Atina Ahdika

Pemodelan Kerugian Bencana Banjir Akibat Curah Hujan Ekstrem Menggunakan EVT dan *Copula*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 13, Nomor 1, Juni 2021, hal 35 – 46

Abstrak

Curah hujan ekstrem pada umumnya dapat mengakibatkan kerugian seperti banjir, tanah longsor, dan gagal panen. Pemodelan kerugian bencana banjir digunakan untuk memperkirakan seberapa parah kerusakan yang dialami ketika terjadi bencana banjir akibat curah hujan ekstrem. Penelitian ini bertujuan untuk memodelkan kerugian bencana banjir terhadap kerusakan rumah menggunakan metode Extreme Value Theory (EVT) dan copula. Metode EVT digunakan untuk memodelkan distribusi curah hujan dan rumah rusak, sedangkan copula digunakan untuk mengidentifikasi struktur dependensi antara curah hujan dengan kerusakan rumah yang ditimbulkan. Hasil dari penelitian ini didapatkan distribusi terbaik untuk kerusakan rumah di Jawa Barat, Jawa Tengah, dan Jawa Timur adalah distribusi Generalized Extreme Value (GEV) serta model copula terbaik yang menggambarkan dependensi antara curah hujan dan kerusakan rumah adalah copula Frank. Nilai parameter copula Frank Provinsi Jawa Barat sebesar 1.4999840280, Jawa Tengah sebesar -0.5816995330, dan Jawa Timur sebesar -0.8648329345. Parameter copula Frank bernilai positif (negatif) menunjukkan hubungan yang sifatnya positif (negatif) antara curah hujan ekstrem dan rumah rusak. Hasil pemodelan menunjukkan bahwa terdapat hubungan positif antara curah hujan dan rumah rusak di Provinsi Jawa Barat dan Jawa Tengah, serta hubungan negatif antara kedua variabel di Provinsi Jawa Timur. Hal tersebut menunjukkan bahwa terdapat

kemungkinan penyebab lain yang lebih berpengaruh terhadap kerusakan rumah di Provinsi Jawa Timur dibandingkan curah hujan.

Kata kunci: Curah Hujan Ekstrem, Banjir, *Extreme Value Theory*, *Copula*

DDC: 315.98

Meinisa Fadillah Rahmi, Paulus Satria Prasetyo, Ratih Nurhabibah, Rizky Perdana, dan Wa Ode Zuhayeni Madjida

Pengelompokan Provinsi di Indonesia Berdasarkan Jumlah Kasus Covid-19 dan Fasilitas Kesehatan

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 13, Nomor 1, Juni 2021, hal 56 – 47

Abstrak

Tingkat penyebaran COVID-19 cukup cepat, hingga 14 November 2020 tercatat jumlah kasus terkonfirmasi positif di Indonesia mencapai 463.007 jiwa. Ketersediaan fasilitas kesehatan masing-masing provinsi menentukan kesiapan daerah dalam penanganan COVID-19 sehingga penting untuk menganalisis keadaan dan distribusi provinsi-provinsi terkait kesiapannya tersebut. Penelitian ini melakukan *clustering* menggunakan algoritma K-Means dan K-Means *with Outlier Detection* untuk mengelompokkan 34 provinsi di Indonesia berdasarkan jumlah kasus COVID-19 dan data fasilitas kesehatan, lalu menentukan metode terbaiknya, serta mengidentifikasi karakteristik masing-masing kelompok berdasarkan metode terbaik. Penelitian menghasilkan tiga *cluster*. *Cluster* 1 merupakan kelompok provinsi dengan jumlah kasus COVID-19 tinggi dan fasilitas kesehatan kurang memadai, *cluster* 2 memiliki jumlah kasus COVID-19 tinggi dan fasilitas kesehatan memadai, sedangkan *cluster* 3 memiliki jumlah kasus COVID-19 rendah dan fasilitas kesehatan menengah.

Kata kunci: COVID-19, *Clustering*, K-Means, K-Means *with Outlier Detection*, Fasilitas Kesehatan



Kata kunci bersumber dari artikel. Lembar abstrak ini boleh diperbanyak tanpa izin dan biaya

DDC: 315.98

Muhamad Saiful Hadi dan Putri Lydia Eltheofany S.

Kajian Usaha Rumah Tangga Budidaya Bandeng di Sulawesi Selatan Tahun 2013

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 13, Nomor 1, Juni 2021, hal 1 – 8

*Abstract*

*From 2010-2013, milkfish production in Indonesia was the highest in the world. The highest production of milkfish in Indonesia is in South Sulawesi. However, the growth of milkfish production in the province from 2010-2012 decreased from 20.67% to 2.75%. This indicates that the milkfish cultivation business actors are still not maximizing the input variable for milkfish production. This study aims to analyze the effect of area, fertilizer and drug costs, and labor costs on milkfish production, to analyze Return to Scale, and to analyze economies of scale in the province of South Sulawesi in 2013. The study used multiple regression analysis methods. Based on the results of the analysis, it is known that land, seed costs, fertilizer and medicine costs, and labor costs have a significant and positive effect with a 95% confidence level on milkfish production in South Sulawesi in 2013. The milkfish cultivation business in South Sulawesi in 2013 was in a state of Rising Back to Scale. The percentage of households that are at economies of scale is 69.53%.*

*Keywords: milkfish cultivation, multiple linear regression, return to scale, economies of scale*

DDC: 315.98

Dadan Kusnandar, Kusnandar, Naomi Nesyana Debataraaja, dan Siti Fitriani

Pemodelan Sebaran *Total Dissolved Solid* Menggunakan Metode *Mixed Geographically Weighted Regression*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 13, Nomor 1, Juni 2021, hal 9 – 16

*Abstract*

*Mixed geographically weighted regression (Mixed GWR) is a technique to explore spatial non-stationarity by considering local and global influences simultaneously. In this study, the Mixed GWR model was used to determine the Total Dissolved Solid (TDS) distribution model on water quality in Pontianak's. Independent variables used are Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD), pH, water hardness and color. The results showed that the Mixed GWR model with Fixed Bisquare Kernel weighting was better than the global regression model and GWR because it had the smallest AIC value of 326.48 and MAPE value of 22.34%. Variables that affect globally are hardness and color, while variables that affect locally are COD, BOD and pH.*

*Keywords: water pollutions, spatial regression, global regression*

DDC: 315.98

Ferra Yanuar, Mutiara Fara Nabilla, dan Izzati Rahmi

Penerapan Peta Kendali Atribut Klasik dan Peta Kendali NP Bayes pada Produk Cacat Air Minum Asri di CV. Multi Rejeki Selaras Payakumbuh

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 13, Nomor 1, Juni 2021, hal 17 – 24

#### *Abstract*

*Efforts that can be made to reduce the number of defective or damaged products is to carry out quality control. Good quality control will help in the smooth production process. One statistical tool that can be used to evaluate whether a production process is in a controlled state or not statistically is the control chart. In this study quality identification is done by creating an attribute control chart, i.e., np control chart. This np control chart is constructed based on the Binomial distribution which will then be assumed using the Bayes method so that the Bayes np control chart is obtained. Next, to test the performance of this control chart, a comparison of the two control charts is performed based on the Average Run Length (ARL) value. The results of the analysis and discussion show that the np Bayes prior control chart conjugate is more sensitive in detecting data out of control than the np Bayes prior non-informative control chart. This control chart also has a better performance because it has a smaller ARL value than the classic np control chart.*

*Keywords: np Control chart, Bayes method, Average Run Length (ARL)*

DDC: 315.98

Nofita Istiana

*Small Area Estimation with Excess Zero (Studi Kasus: Angka Kematian Bayi di Pulau Jawa)*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 13, Nomor 1, Juni 2021, hal 25 – 34

#### *Abstract*

*Infant Mortality Rate (IMR) is an important indicator because it can be used to compare health status between populations. IMR is obtained from Demographic and Health Survey (DHS) where the level of estimation is designed for national and provincial level. The decentralization system makes the importance of IMR for sub-domain of province such as district/municipality level. Small area estimation (SAE) can be used for estimating IMR in district/municipality level by using a mixed model. IMR is count data with small probability, so the distribution is Poisson. Poisson model assumes that mean equal to variance, but this assumption is often violated. One of the reasons is excess zero. In this study, Zero Inflated Poisson (ZIP) mixed model is much better than Poisson mixed model and can improve the direct estimation of IMR.*

*Keywords: SAE, IMR, excess zero, count data, ZIP mixed model*

Ian Surya Prayoga dan Atina Ahdika

Pemodelan Kerugian Bencana Banjir Akibat Curah Hujan Ekstrem Menggunakan EVT dan *Copula*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 13, Nomor 1, Juni 2021, hal 35 – 46

*Abstract*

*In general, extreme rainfall can result in losses such as floods, landslides, and crop failure. Flood loss modeling is used to estimate how much damage is experienced when a flood occurs due to extreme rainfall. This study aims to model the losses of floods on house damage using the Extreme Value Theory (EVT) and copula methods. The EVT method is used to model the distribution of rainfall and damaged houses, while the copula is used to identify the dependency structure between the rainfall and the damage to the houses it causes. The results of this study show that the best distribution for house damage in West Java, Central Java and East Java is the Generalized Extreme Value (GEV) distribution and the best copula model that describes the dependency between rainfall and house damage is Frank copula. The value of the Frank copula parameter in West Java Province is 1.4999840280, in Central Java is -0.5816995330, and in East Java is -0.8648329345. The positive (negative) value of Frank copula parameter indicating a positive (negative) relationship between extreme rainfall and damaged houses. The modeling results show that there is a positive relationship between rainfall and damaged houses in West Java and Central Java Provinces, as well as a negative relationship between the two variables in East Java Province. This indicates that there are other possible causes that are more influencing the damage to houses in East Java Province than rainfall.*

DDC: 315.98

Meinisa Fadillah Rahmi, Paulus Satria Prasetyo, Ratih Nurhabibah, Rizky Perdana, dan Wa Ode Zuhayeni Madjida

Pengelompokkan Provinsi di Indonesia Berdasarkan Jumlah Kasus Covid-19 dan Fasilitas Kesehatan

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 13, Nomor 1, Juni 2021, hal 56 – 47

*Abstract*

*The rate spread of COVID-19 is quite fast, until November 14, 2020, the number of cases in Indonesia has reached 463,007 people. The availability of health facilities for each province determines regional readiness in handling COVID-19, so it is very important to analyze the condition and distribution of provinces related to their readiness. This study conducted clustering using K-Means and K-Means with Outlier Detection algorithm to classify 34 provinces in Indonesia based on the number of COVID-19 cases and health facilities, then decide the best model, and identify the characteristics of each group based on the best model. The study resulted in three clusters. Cluster 1 is a group of provinces with a high number of COVID-19 cases and inadequate health facilities, cluster 2 has a high number of COVID-19 cases and adequate health facilities, while cluster 3 has a low number of COVID-19 cases and intermediate health facilities.*

*Keywords: COVID-19, Clustering, K-Means, K-Means with Outlier Detection, Health Facilities*

# **PENERAPAN *SOCIAL NETWORK ANALYSIS* DAN *LATENT DIRICHLET ALLOCATION* UNTUK PEMETAAN PUBLIKASI PENELITIAN DOSEN POLITEKNIK STATISTIKA STIS**

**Muhammad Iqbal<sup>1</sup> dan Setia Pramana<sup>1,2\*</sup>**

<sup>1</sup>Politeknik Statistika STIS, <sup>2</sup> Direktorat Analisis dan Pengembangan Statistik, Badan Pusat Statistik  
\*e-mail: setia.pramana@stis.ac.id

## **Abstrak**

Penelitian merupakan salah satu tugas utama institusi pendidikan, dimana dosen adalah motor dalam pengembangan penelitian. Pemetaan dosen berdasarkan pola kolaborasi serta topik yang dikaji dibutuhkan sebagai salah satu dasar untuk pengembangan kualitas dan kuantitas penelitian. Penelitian ini bertujuan untuk menganalisis pola kolaborasi antar dosen serta topik utama dari hasil penelitian dosen Politeknik Statistika STIS. Analisis penelitian ini menggunakan metode *Social Network Analysis* (SNA) dan *Latent Dirichlet Allocation* (LDA). Dari hasil analisis terlihat pola hubungan kolaborasi penelitian antar dosen serta dosen dengan posisi sentral pada proses kolaborasi tersebut. Kemudian dari hasil LDA terlihat beberapa topik-topik utama dari publikasi ilmiah seperti implementasi analisis regresi, kemiskinan, serta analisis data panel dalam bidang ekonomi.

**Kata kunci:** Publikasi ilmiah, *Social Network Analysis*, *Latent Dirichlet Allocation*

## **Abstract**

*Research is one of main function of an educational institution where lecturer plays important roles. Lecture mapping based on their collaboration and research topic is needed for the institution and stakeholders to improve the quantity and quality of the research. This study aims to analyze the scientific publications of Politeknik Statistika STIS lecturers to see lecturer collaboration and the themes of the research. The analysis uses the Social Network Analysis (SNA) and Latent Dirichlet Allocation (LDA) methods. The results show a network graphic illustrating how research collaboration relationship between lecturers and also identify lectures with central position in the collaboration. The LDA graph shows several main topics from scientific publications such as regression analysis, poverty, and panel data analysis for economics.*

**Keywords:** *Scientific publications, Social Network Analysis, Latent Dirichlet Allocation*

## PENDAHULUAN

Pendidikan Tinggi adalah jenjang pendidikan setelah pendidikan menengah yang mencakup program diploma, program sarjana, program magister, program doktor, dan program profesi, serta program spesialis, yang diselenggarakan oleh perguruan tinggi berdasarkan kebudayaan bangsa Indonesia. Satuan pendidikan yang melaksanakan pendidikan tinggi adalah Perguruan Tinggi. Berdasarkan Tridharma Perguruan Tinggi, tugas dosen sebagai Sivitas Akademika ada tiga, yaitu pendidikan, penelitian, dan pengabdian kepada masyarakat. Pembelajaran dan penelitian merupakan pilar utama dalam pendidikan tinggi. Pemerintah Indonesia telah mengeluarkan sejumlah kebijakan peraturan untuk meningkatkan minat para peneliti dalam mempublikasikan hasil penelitiannya, baik secara nasional maupun internasional. Melalui Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 17 tahun 2013 tentang Jabatan Fungsional Dosen dan Angka Kreditnya, pemerintah mewajibkan dosen yang ingin memperoleh jabatan akademik Asisten Ahli, atau kenaikan jabatan dari Asisten Ahli ke Lektor, atau dari Lektor Kepala untuk melakukan penelitian dan publikasi ilmiah.

Pemetaan dosen berdasarkan hasil penelitiannya baik dari sisi topik serta bagaimana berkolaborasi antar dosen sangat diperlukan untuk meningkatkan kualitas serta kuantitas penelitian di suatu institusi pendidikan. Pemetaan ini sendiri masih belum tersedia di Politeknik Statistika STIS dan dibutuhkan oleh Pusat Penelitian dan Pengabdian Masyarakat (P3M) Politeknik Statistika STIS untuk mengelompokkan dosen berdasarkan kolaborasi penelitiannya dan memetakan kompetensi dosen berdasarkan konten tulisan, sehingga penelitian ini berfokus pada analisis kolaborasi penelitian dan tema dari penelitian yang dilakukan dosen Politeknik Statistika STIS. Metode yang digunakan untuk menentukan kolaborasi dosen dan tema dari penelitian tersebut adalah *Social*

*Network Analysis* (SNA) dan *Latent Dirichlet Allocation* (LDA).

*Social Network Analysis* (SNA) adalah satu alat untuk memetakan hubungan pengetahuan penting antara individu (Pryke, 2004). Dengan menggunakan SNA terhadap hubungan sosial dapat diketahui ukuran hubungan antar aktor dalam hubungan tersebut, seperti seberapa sering mereka berinteraksi, siapa yang menjadi pusat dari hubungan tersebut, dan lain-lain. *Latent Dirichlet Allocation* (LDA) menurut David M. Blei, dkk (2003) adalah model probabilistik generatif dari korpus. LDA memungkinkan set pengamatan bisa dijelaskan oleh kelompok yang tidak teramati untuk menjelaskan kemiripan beberapa bagian data. Penggunaan model LDA memungkinkan untuk mendapatkan topik-topik umum dari suatu korpus dokumen.

Dengan menggunakan metode SNA pada publikasi ilmiah dosen bisa dihasilkan sebuah grafik jaringan yang memvisualisasikan kolaborasi penelitian antar dosen dan menampilkannya datanya secara informatif, dan didapatkan juga ukuran sentralistis dosen untuk mengetahui pengaruh dosen pada jaringan kolaborasi penelitian. Dan dengan menggunakan metode LDA pada publikasi ilmiah dosen akan didapatkan topik-topik umum dari publikasi ilmiah dosen yang dapat digunakan untuk mengelompokkan publikasi berdasarkan topik-topik tersebut. Hasil penelitian ini diharapkan bermanfaat untuk mempermudah P3M Politeknik Statistika STIS untuk memetakan kolaborasi dosen dan tema penelitian.

Dengan meningkatnya jumlah publikasi ilmiah dosen Politeknik Statistika STIS dari tahun ke tahun menyebabkan tidak memungkinkan untuk melakukan pemetaan dosen berdasarkan publikasi ilmiah secara manual. Pemetaan dosen berdasarkan kolaborasi penelitian dan tema publikasi ilmiah dengan tanpa pengawasan (*unsupervised*) masih belum dilakukan di Politeknik Statistika STIS. Tujuan dari penelitian ini adalah untuk menganalisis kolaborasi dosen dan tema publikasi tanpa

perlu adanya pengawasan menggunakan metode SNA dan LDA.

Terdapat beberapa penelitian terkait dengan penelitian ini. Penelitian terkait pertama adalah penelitian yang dilakukan oleh Hennie Tuhuteru dan Ade Iriani (2018) dengan judul “Analisis Kolaborasi Penelitian Dosen Fakultas X dengan Social Network Analysis (SNA)”. Penelitian terkait selanjutnya adalah penelitian yang dilakukan oleh I Made Kusnanta Bramantya Putra (2017) dengan judul “Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA)”.

Pada penelitian Henni (2018), dosen divisualisasikan berdasarkan jenis kelamin, *node* pria berbentuk lingkaran dan *node* perempuan berbentuk kotak. Pewarnaan *node* dilakukan berdasarkan atribut program studi dari masing-masing aktor. Dengan penghitungan didapatkan nilai *degree centrality*, *closeness centrality*, dan *betweenness centrality* dengan kesimpulan yang didapat adalah jurusan yang sering melakukan kolaborasi, jumlah kolaborasi, dan hubungan Jabatan Fungsional Akademik yang sering melakukan kolaborasi.

Sedangkan Putra (2017) melakukan penelitian pada pendapat publik terhadap stasiun radio dengan mengklasifikasikan pendapat-pendapat tersebut ke dalam beberapa topik. Dengan jumlah topik yang didapatkan adalah 4, diuji dengan mesin didapatkan nilai perplexity sebesar 231,4 dan diuji kemudahannya untuk diinterpretasi oleh manusia melalui uji koherensi topik yang terdiri dari *Word Intrusion Task* dan *Topic Intrusion Task*. Kesimpulan dari uji koherensi topik menyatakan bahwa model yang dihasilkan dengan metode LDA pada studi kasus ini dapat diinterpretasi manusia dengan baik.

## METODOLOGI

### 1. Tinjauan Referensi

#### Social Network Analysis (SNA)

Menurut Lin Zhu, dkk (2019) *Social Network Analysis* (SNA) adalah sebuah proses yang digunakan untuk memeriksa

struktur sosial dengan menggunakan teori jaringan dan grafik. Freeman (1989) menyatakan bahwa minimal basis data jaringan terdiri dari satu set objek (aktor atau *node*) yang dihubungkan oleh suatu hubungan pada suatu kejadian (*edge*). Representasi matriks yang umum digunakan untuk menggambarkan jaringan adalah matriks sosiometri atau matriks “dari siapa ke siapa”. Jika jaringan sosial dimisalkan sebagai sktruktur sosial yang terdiri dari individu ataupun organisasi, Wasserman & Faust (1994) mengatakan bahwa *nodes* jaringan tersebut adalah individu atau organisasi tersebut, dimana *edgenya* bisa berupa hubungan seperti persahabatan, kekerabatan, kepentingan bersama, pertukaran uang, tidak sukaan, dan pengetahuan.

Berdasarkan Carrington, dkk (2005) SNA sudah digunakan sejak pertengahan tahun 1930-an dan mengalami berbagai macam perkembangan mencakup sosiometri, teori grafik, *dyads*, *triads*, *subsgroup*, blok model dan masih terus berkembang sampai tahun 2000-an. Ada berbagai macam metode untuk mengukur jaringan sosial, salah satunya metodenya adalah yang dikemukakan oleh Freeman (1979) yaitu dengan menggunakan metode analisis sentralistis. ia mengembangkan tiga konsep utama analisis sentralistis, yaitu *degree*, *closeness*, dan *betweenness*. Metode analisis sentralistis memungkinkan untuk menganalisis sentralistis berbagai jenis macam data, dari *one-mode* hingga *two-mode data*.

Pengukuran sentralistis yang digunakan adalah mengukur tiga kategori dasar sentralistis, yaitu *degree*, *closeness*, dan *betweenness*. Meskipun sudah banyak analisis sentralistis yang telah ditemukan, tiga ini mendominasi penggunaan empiris.

#### 1. Degree Centrality

Shaw (1954) memperkenalkan ide untuk menggunakan *degree* sebagai indeks sentralitas. Misal pada jaringan komunikasi, seseorang yang berada pada posisi yang melakukan kontak langsung dengan orang banyak, maka orang tersebut bisa dianggap sebagai saluran informasi utama atau dengan kata lain titik fokus dari

komunikasi. Jadi *degree* melihat sentralitas berdasarkan banyaknya aktivitas yang dilakukan.

Nieminen (1974) merumuskan cara menghitung *degree* ( $C_D$ ) suatu titik  $p_k$ :

$C_D(p_k) = \sum_{i=1}^n a(p_i, p_k)$ , dimana  $a(p_i, p_k) = 1$  jika dan hanya jika  $p_i$  dan  $p_k$  terhubung oleh garis, 0 jika tidak.

Semakin banyak jumlah titik yang terhubung maka semakin besar nilai *degree*, *degree* sama dengan 0 jika  $p_k$  tidak terhubung dengan titik mana pun, jadi rumus ini berguna untuk menghitung jumlah aktivitas yang terhubung dari titik tersebut. Dengan menggunakan rumus *degree* bisa didapatkan *degree centrality* dengan:

$C_D'(p_k) = \frac{\sum_{i=1}^n a(p_i, p_k)}{n-1}$ , dimana  $n-1$  adalah nilai maksimal dari  $C_D(p_k)$ .

## 2. Betweenness Centrality

Bavelas (1948) dan Shaw (1954) mengemukakan bahwa ketika seseorang secara strategis berada di jalur yang menghubungkan sepasang orang lain, maka orang yang terletak pada posisi tersebut merupakan adalah pusat dari jalur tersebut. Misal pada jaringan komunikasi, orang yang berposisi sebagai penghubung komunikasi antara dua orang merupakan pusat karna orang tersebut bisa mengontrol komunikasi mereka. Jadi *betweenness* melihat sentralitas berdasarkan posisi pada jalur antara dua atau lebih *node* pada jaringan.

Cara mengukur *betweenness* dikemukakan oleh Anthonisse (1971) dan Freeman (1977) dengan cara menghitung jarak antara dua titik, dan melakukan perhitungan peluang jika jalurnya lebih dari satu. Rumus untuk menghitung *betweenness* yang dikemukakan:

$$B_{ij}(P_k) = \frac{g_{ij}(p_k)}{g_{ij}}$$

dimana

$g_{ij}$  = banyak jalur terpendek antara titik  $p_i$  dan  $p_j$

$g_{ij}(p_k)$  = banyaknya jalur terpendek antara titik  $p_i$  dan  $p_j$  yang mengandung titik  $p_k$

dan rumus untuk menghitung *betweenness* secara keseluruhan:

$$C_B(p_k) = \sum_{i=1}^n \sum_{j=1}^n B_{ij}(p_k),$$

dimana  $i < j$ ,  $i \neq j \neq k$ , dan  $n$  banyaknya titik pada grafik. Setiap kali  $p_k$  menjadi titik yang hanya menghubungkan dua pasang titik maka nilai  $C_B(p_k)$  bertambah 1.

Freeman (1977) menunjukkan bahwa rumus *betweenness* secara keseluruhan adalah:

$$C_B'(p_k) = \frac{2C_B(p_k)}{n^2 - 3n + 2}$$

Rumus  $C_B'(p_k)$  pada suatu titik akan menghasilkan nilai 1 jika semua titik yang ada hanya terhubung ke titik tersebut, sehingga pada grafik itu  $C_B'(p_k)$  akan menghasilkan nilai 0 jika dicoba pada titik yang lain.

## 3. Closeness Centrality

*Closeness* berkaitan dengan kontrol komunikasi untuk melihat sejauh mana pengaruh dari orang lain dapat dihindari, atau kata lainnya independensi. Berdasarkan Bavelas (1950), posisi non-sentral adalah orang yang harus menyampaikan informasinya melalui orang lain. Leavit (1951) berpendapat bahwa orang dalam posisi pusat dalam contoh kasus penyampaian informasi tidak terlalu bergantung ke orang lain dalam menyampaikan informasi, bahkan menurut Leavit kata sentralistis dan independen bisa ditukar.

Berbagai peneliti mencoba menjelaskan *closeness*, Bavelas (1948) menjelaskan bahwa informasi yang berasal dari titik pusat membutuhkan waktu yang minimal disampaikan ke seluruh titik pada jaringan. Beuchamp (1965) membandingkannya dengan organisasi seperti efisiensi, optimal, dll. dalam melakukan komunikasi. Hakimi (1965) dan Sabidussi (1966) mendefinisikannya secara umum yaitu titik tengah dalam melakukan komunikasi ke semua titik menggunakan biaya dan waktu yang paling minimum. Jadi secara berdasarkan uraian tersebut *closeness* adalah titik yang mempunyai jarak yang paling minimum untuk mencapai semua titik.

Sabidussi (1966) merumuskan cara menghitung *closeness* dengan cara menjumlahkan semua jarak terpendek yang menghubungkan suatu titik ke titik yang



lain. Rumus ini untuk mengukur desentralisasi titik atau invers *closeness*:  $C_C(p_k)^{-1} = \sum_{i=1}^n d(p_i, p_k)$ , dimana  $d(p_i, p_k)$  = banyaknya jalur terpendek yang menghubungkan antara  $p_i$  dan  $p_k$ .

Semakin banyak jarak antara  $p_i$  dan  $p_k$  maka nilai  $C_C(p_k)^{-1}$  semakin besar.  $C_C(p_k)^{-1}$  akan bernilai tak hingga jika digunakan pada titik-titik yang tidak terhubung.

Beauchamp (1965) merumuskan *closeness* secara keseluruhan dengan rumus:

$$C'_C(p_k) = \frac{n - 1}{\sum_{i=1}^n d(p_i, p_k)}$$

Pengaplikasian ketiga sentralistis tersebut tergantung dengan konteks aplikasi yang digunakan. Jika berkaitan dengan aktivitas komunikasi bisa dengan menggunakan pengukuran *degree*. Jika berkaitan dengan kontrol komunikasi bisa dengan menggunakan pengukuran *betweenness*. Jika berkaitan dengan independensi dan efisiensi bisa dengan menggunakan pengukuran *closeness*.

### **Latent Dirichlet Allocation (LDA)**

Baeza-Yates dan Ribeiro-Neto (1999) mengusulkan metodologi dasar untuk pengambilan informasi dari korpus teks, sebuah metodologi yang sering digunakan dalam mesin pencarian, yaitu mereduksi setiap dokumen dalam korpus menjadi vektor bilangan real, yang masing-masing mewakili rasio jumlah dari kata.

Dalam skema *term frequency – invers document frequency* (tf – idf) (Salton dan McGill, 1983), kosaka yang digunakan untuk melambangkan unit dasar dari korpus adalah “kata” atau “istilah”, dimana hitungannya dibentuk dari jumlah kemunculannya untuk setiap dokumen pada korpus. Setelah dinormalisasi, jumlah kata ini dibandingkan dengan invers jumlah dokumen, yang mengukur jumlah kemunculan kata di seluruh korpus. Hasil akhirnya adalah matrik X kata per dokumen yang kolomnya terdiri dari nilai tf-idf untuk masing-masing dokumen dalam korpus. Jadi dengan menggunakan skema tf-idf dokumen yang panjang mengecil menjadi daftar angka.

Meski skema tf-idf mempunyai beberapa kelebihan, pendekatan ini juga memiliki beberapa kelemahan, yaitu pendekatan ini hanya menyediakan relatif kecil pengurangan dan mengungkapkan sedikit dari struktur statistik antar/dalam dokumen. Untuk mengatasi kekurangan ini beberapa peneliti telah mengusulkan beberapa metode pengurangan dimensionalitas lainnya, dan kebanyakan metode ini berdasarkan pada asumsi “*bag of word*”, yaitu urutan dari kata dalam dokumen diabaikan. Dan meski kurang dinyatakan secara formal, metode ini juga mengasumsikan urutan dokumen dalam korpus dapat diabaikan. Asumsi dapat dipertukarkan tidak berarti sama dengan asumsi variabel acak yang berdistribusi independen dan identik.

De Finetti (1990) menetapkan bahwa setiap kumpulan variabel acak yang dapat dipertukarkan memiliki representasi sebagai distribusi campuran, yang secara umum campuran yang tidak terhingga. Jadi jika ingin mempertimbangkan representasi untuk dokumen dan kata yang dapat dipertukarkan maka perlu dipertimbangkan model campuran yang dapat menangkap pertukaran kata dan dokumen tersebut, alasan inilah terbentuknya *Latent Dirichlet*, untuk menangkap struktur statistik intra-dokumen yang signifikan melalui distribusi campuran.

Istilah-istilah yang digunakan berupa:

- Kata merupakan unit dasar dari data diskrit, yang didefinisikan sebagai item dari kosakata yang diindeks dengan  $\{1, \dots, V\}$ . Kata-kata diwakili oleh vektor satuan dasar yang memiliki satuan tunggal sama dengan satu dan semua komponen lainnya sama dengan nol.
- Dokumen adalah urutan dari  $N$  kata-kata yang dinotasikan dengan  $w = (w_1, w_2, \dots, w_N)$ , dimana  $w_N$  adalah kata ke  $n$  dalam urutan.
- Korpus adalah kumpulan  $M$  dokumen yang dinotasikan dengan  $D = \{w_1, w_2, \dots, w_M\}$

Blei dkk (2003) menjelaskan LDA sebagai model probabilistik Bayesian tiga tingkat untuk sekumpulan data diskrit seperti korpus. Ide dasarnya adalah

dokumen merupakan campuran acak dari serangkaian topik yang tidak kelihatan, dimana setiap topik dibentuk dari distribusi berbagai kata tertentu.

LDA mengasumsikan proses generatif berikut untuk setiap  $w$  dokumen dengan  $D$  korpus:

1. Pilih  $N \sim \text{Poisson}(\xi)$ .
2. Pilih  $\theta \sim \text{Dir}(\alpha)$ .
3. Untuk setiap  $N$  kata  $w_n$ :
  - a. Pilih topik  $z_n \sim \text{Multinomial}(\theta)$ .
  - b. Pilih kata  $w_n$  dari  $p(w_n | z_n, \beta)$ , probabilitas multinomial yang dikondisikan pada topik  $z_n$ .

Simplistiknya rumus dilakukan dengan menggunakan asumsi:

1. Dimensionalitas  $k$  dari distribusi dirichlet dan dimensionalitas dari topik variabel  $z$  diketahui dan tetap.
2. Peluang dari kata diparameterkan dengan  $k \times V$  matrix  $\beta$  yang diperlakukan sebagai jumlah yang tetap untuk diestimasi.
3. Asumsi Poisson dianggap tidak terlalu penting, karena  $N$  tersebut juga independen dari variabel penghasil data ( $\theta$  dan  $z$ ).

Dengan menggunakan parameter  $\alpha$  dan  $\beta$ , maka distribusi gabungan dari campuran topik  $\phi$ , satu set  $N$  topik  $Z$ , dan satu set kata  $W$  adalah:

$$p(\theta, z, w | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(\theta) p(w_n | z_n, \beta),$$

dimana  $p(z_n | \theta)$  sama dengan  $\theta_i$  untuk  $i$  unik seperti  $z_n^i = 1$ . Dengan mengintegrasikan  $\theta$  dan menjumlahkan  $z$ , maka didapatkan distribusi marginal dari dokumen:

$$p(w | \alpha, \beta) = \int p(\theta | \alpha) \left( \prod_{n=1}^N \sum_{z_n} p(\theta) p(w_n | z_n, \beta) \right) d\theta$$

Dengan mengambil hasil dari probabilitas marginal dari dokumen tunggal, maka didapatkan peluang korpus:

$$p(D | \alpha, \beta) = \prod_{d=1}^M \int p(\theta_d | \alpha) \left( \prod_{n=1}^{N_d} \sum_{z_{dn}} p(\theta_d) p(w_{dn} | z_{dn}, \beta) \right) d\theta_d$$

## Perplexity

*Perplexity* adalah ukuran seberapa baik model probabilitas dalam menerima set data yang baru. Blei, dkk. (2003) menggunakan *perplexity* dari set uji untuk mengevaluasi model. Nilai *perplexity* secara monoton menurun dari kemungkinan uji data. Semakin rendah nilai *perplexity* menunjukkan kinerja generalisasi yang lebih baik. Berdasarkan penelitian tersebut, dengan  $M$  dokumen uji, *perplexity* adalah:

$$\text{Perplexity}(D_{\text{test}}) = \exp \left\{ \frac{\sum_{d=1}^M \log p(w_d)}{\sum_{d=1}^M N_d} \right\}.$$

## Coherence

Sekumpulan pernyataan atau fakta dikatakan berkoherensi jika sekumpulan pernyataan atau fakta tersebut saling mendukung. Dalam filsafat ilmiah, pendekatan yang digunakan adalah dengan menggunakan gabungan fungsi dan probabilitas marginal yang terasosiasi dengan faktanya. Penggunaan koherensi pada topik model muncul karena pada topik model belum adanya jaminan interpretabilitas pada hasil topik model tersebut. Newman, dkk. (2010) mengusulkan langkah-langkah koherensi otomatis yang menilai topik terkait dengan pemahaman pada topik tersebut. Caranya dengan memperlakukan kata-kata pada topik sebagai fakta terus membatasi koherensi yang digunakan didasarkan pada perbandingan sepasang kata. Evaluasi pada penelitian tersebut adalah peringkat topik diukur berdasarkan pada co-occurrence kata.

Aletra dan Stevenson (2013) memperkenalkan koherensi topik yang berdasar pada vektor konteks untuk setiap kata teratas pada topik. Vektor konteks dari kata  $w$  dibuat dengan menghitung co-occurrence kata yang ditentukan menggunakan jendela konteks yang berisi semua kata yang terletak kurang lebih 5 token di sekitar kemunculan dari kata  $w$ . Element vektor didefinisikan menggunakan Normalized Pointwise Mutual Information (NPMI) untuk mendapatkan korelasi terbaik. Sehingga element ke- $j$  pada vektor konteks  $v_i$  dari kata  $w_i$  mempunyai NPMI:

$$v_{ij} = NPMI(w_i, w_j)^\gamma$$

$$= \left( \frac{\log \frac{P(w_i, w_j) + \epsilon}{P(w_i) + P(w_j)}}{-\log(P(w_i, w_j) + \epsilon)} \right)^\gamma$$

Berdasarkan Roder, dkk. (2015) topik koherensi menggunakan vektor memiliki rata-rata hasil yang lebih baik dibandingkan topik koherensi yang lain.

## Relevance

Bischof dan Ailordi (2012) mengajukan sebuah metode untuk memberikan peringkat terhadap istilah dari suatu topik tertentu baik dari segi frekuensi istilah di topik tersebut maupun eksklusivitas istilah tersebut pada topik, dengan memperhitungkan seberapa banyak istilah tersebut muncul pada topik tersebut dengan mengesampingkan topik lainnya. Siever dan Shirley (2014) mengajukan sebuah pengukuran yang sama dengan nama Relevance yang memungkinkan pengguna untuk menentukan peringkat istilah secara fleksibel untuk memudahkan pengguna dalam menafsirkan sebuah topik. Relevance diterapkan pada LDA karena secara umum topik yang dihasilkan susah untuk ditafsirkan manusia.

Berdasarkan penelitian Siever dan Shirley (2014), misalkan  $\phi_{wk}$  menotasikan peluang istilah  $w \in \{1, \dots, V\}$  untuk topik  $k \in \{1, \dots, K\}$ , dimana  $V$  menotasikan banyak istilah dalam kosakata, dan misalkan  $p_w$  menotasikan peluang marginal dari istilah  $w$  pada korpus. Rumus Relevance pada istilah  $w$  pada topik  $k$  dengan parameter bobot  $\lambda$  (dimana  $0 \leq \lambda \leq 1$ ) adalah:

$$Y(w, k | \lambda) = \lambda \log \log (\phi_{kw}) + (1 - \lambda) \log \left( \frac{\phi_{wk}}{p_w} \right),$$

dimana  $\lambda$  adalah bobot yang diberikan pada probabilitas suatu istilah  $w$  dalam suatu topik  $k$  relatif terhadap kenaikannya. Ketika  $\lambda = 1$  maka istilah diberikan peringkat berdasarkan peluang istilah tersebut dalam topik, sedangkan ketika  $\lambda = 0$  maka istilah tersebut hanya diperingkat berdasarkan liftnya, yaitu perbandingan antara probabilitas istilah pada topik

tersebut dengan probabilitas marginal istilah pada korpus. Jadi ketika  $\lambda = 1$  maka istilah yang umum pada topik peringkatnya akan tinggi, dan ketika  $\lambda = 0$  maka istilah yang spesifik pada topik akan tinggi. Relevansi yang digunakan pada penelitian ini adalah 0,6.

## 2. Metode Analisis

Data yang digunakan dalam penelitian adalah data teks dari situs repositori Sistem Informasi Terpadu (SIPADU) Politeknik Statistika STIS. SIPADU merupakan sistem informasi yang dikembangkan dan diimplementasikan secara mandiri oleh para alumni STIS untuk mendukung kegiatan seluruh civitas akademik di STIS. Data yang diambil berupa nama, penulis, dan judul. Data tersebut dikumpulkan dengan metode *web scraping* menggunakan bahasa pemrograman Python dengan *package selenium* untuk mengambil elemen-elemen yang terdapat pada suatu halaman situs, sehingga dengan mengombinasikannya dengan algoritma pemrograman maka pemrograman bisa mengambil, mengolah, dan menyimpan data-data dari halaman repositori. Pendeteksian elemen pada situs menggunakan *full Xpath* agar bisa dilakukan algoritma pengulangan untuk mengambil data satu per satu seluruh data yang dibutuhkan. Data yang telah dikumpulkan sebanyak 652 data publikasi dengan data nama unik dosen penulis yang tersedia sebanyak 64 dosen Politeknik Statistika STIS. Dikarenakan tidak tersedianya data abstrak dan kata kunci untuk diambil menggunakan *scrapping* maka data yang digunakan untuk metode LDA hanya data judul publikasi, yang mungkin akan menyebabkan hasil yang didapat tidak sesuai dengan hasil yang diharapkan karena kekurangan data untuk diolah.

Data judul publikasi dan data nama penulis yang telah dikumpulkan kemudian dilakukan *preprocessing* seperti penghapusan tanda kurung, *case folding*, *remove punctuation*, *remove number*, *tokenization*, *remove stopword*, dan *stemming*. Untuk data nama penulis

dilakukan penghapusan tanda kurung, *case folding*, *remove punctuation*, *remove number*, *tokenization*, menghapus gelar menggunakan *stopword*, mengganti kata ‘dan’ dan ‘and’, dan terakhir dibuat variabel *source* menggunakan nama unik penulis dan variabel *target* berupa data nama yang telah dilakukan *preprocessing* untuk membentuk variabel baru yang dinamakan Concat yang merupakan gabungan dari dua variabel tersebut. Data yang telah diolah disimpan ke tabel DBMS MySQL menggunakan package *PyMySQL*.

Untuk melakukan SNA digunakan package *networkx* metode *adjacency list* pada variabel Concat akan menghasilkan data dalam bentuk SNA, yaitu tersedianya *node*, *edge*, *degree*, data ini dinotasikan dengan simbol G. Karena G masih mengandung nama-nama dosen di luar Politeknik Statistika STIS maka nama dosen selain dari PolStat STIS dikeluarkan. Variable dari G yang digunakan adalah *edge*, *degree*, *degree centrality*, *betweenness centrality*, dan *closeness centrality*. Untuk keperluan visualisasi agar lebih informatif dilakukan penghitungan per nama sebagai informasi jumlah publikasi.

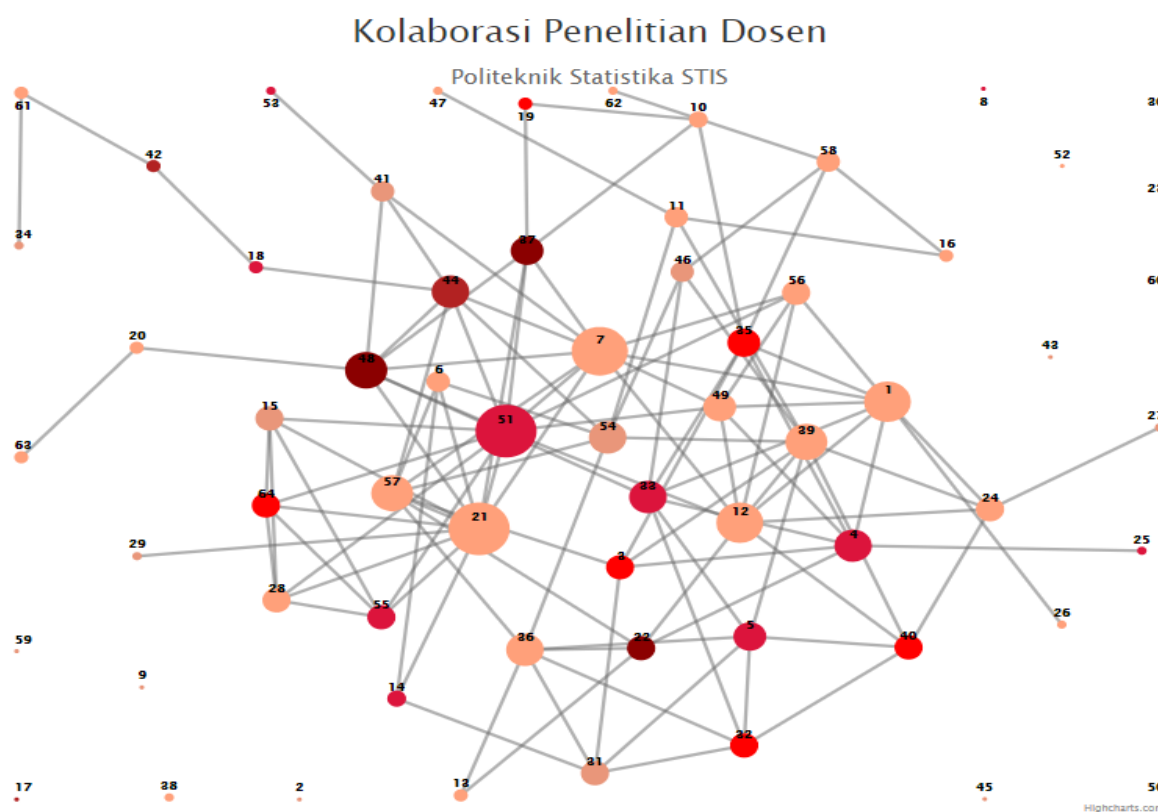
Kemudian LDA diimplementasikan pada data judul publikasi. Dari data judul publikasi tersebut dilakukan pembentukan *dictionary* dan *corpus* menggunakan package *gensim*. *Dictionary* dan *corpus* akan digunakan untuk membentuk model LDA dengan jumlah topik yang ditentukan. Package *pyLDAvis* digunakan pada model untuk menghasilkan visualisasi grafik LDA.

Untuk menguji ketepatan model SNA dilihat dari ukuran sentralitasnya berupa *degree centrality*, *betweenness centrality*, dan *closeness centrality*. Visualisasi SNA dilakukan dengan menggunakan *JavaScripts Highcharts* untuk menghasilkan *network graph*.

Pengujian terhadap model LDA dilakukan dengan menentukan jumlah topik yang tepat untuk model. Pengujian ini dilakukan dengan menggunakan uji *perplexity* dan uji *coherence*. Jumlah topik yang baik adalah jumlah topik yang memiliki nilai uji yang tinggi dan efisien.

## HASIL DAN PEMBAHASAN

### Implementasi SNA

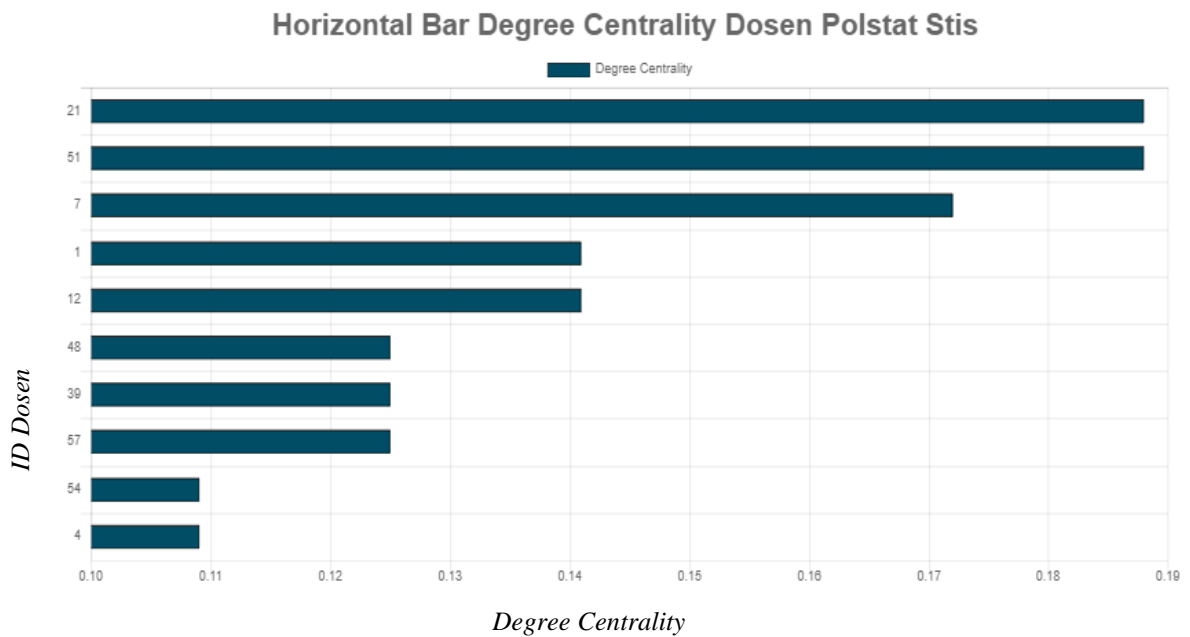


Gambar 1. Grafik jaringan kolaborasi penelitian dosen Politeknik Statistika STIS

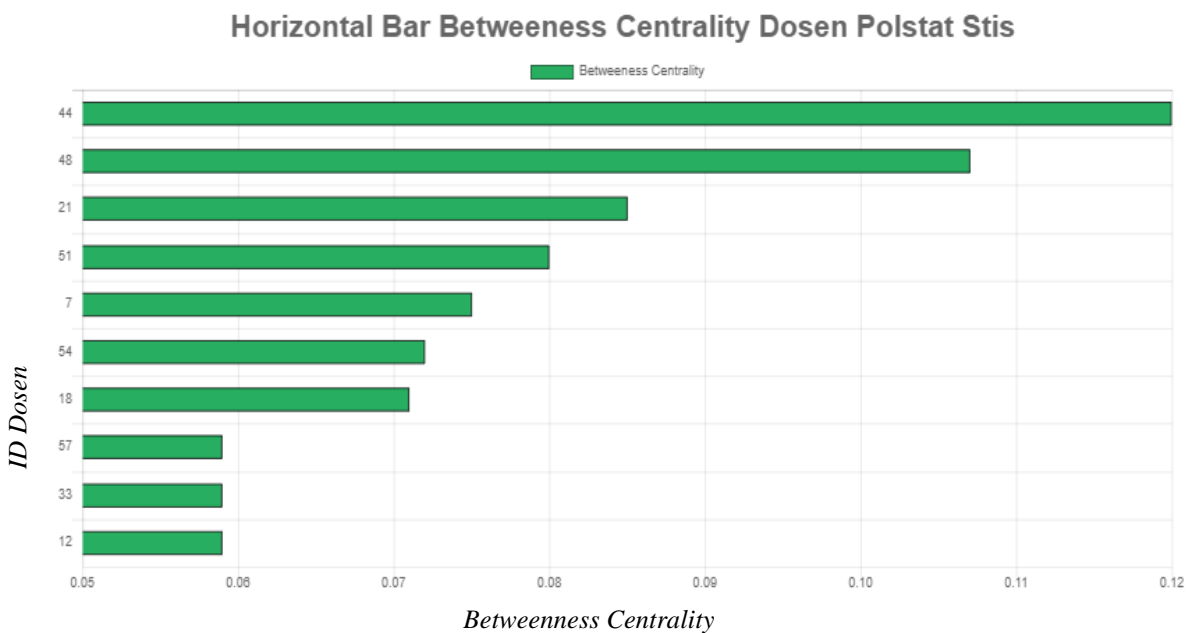
Dari hasil scraping didapat 652 publikasi ilmiah yang terambil dari SIPADU PolStat STIS. Implementasi SNA menggunakan package Python networkx untuk menghasilkan data dalam bentuk *node* dan *edge* menggunakan data dosen Politeknik STIS dari database yang telah melaporkan publikasinya. Data yang digunakan adalah nama dan penulis yang telah dilakukan preprocessing. Visualisasi SNA pada website menggunakan JavaScripts Highcharts untuk menghasilkan network graph seperti pada gambar 1.

Model SNA yang dihasilkan ditampilkan dalam Gambar 1. *Node* menunjukkan data dosen Politeknik Statistika STIS yang telah melaporkan publikasinya dan *edge* menunjukkan kolaborasi penelitian dengan dosen Politeknik Statistika STIS lainnya.

Semakin gelap warna pada *node* maka akan semakin tinggi rentang jumlah publikasinya. Ukuran lingkaran pada *node* menunjukkan seberapa banyak dosen yang pernah diikuti sertakan dalam kolaborasi penelitian. Terlihat bahwa dosen dengan ID

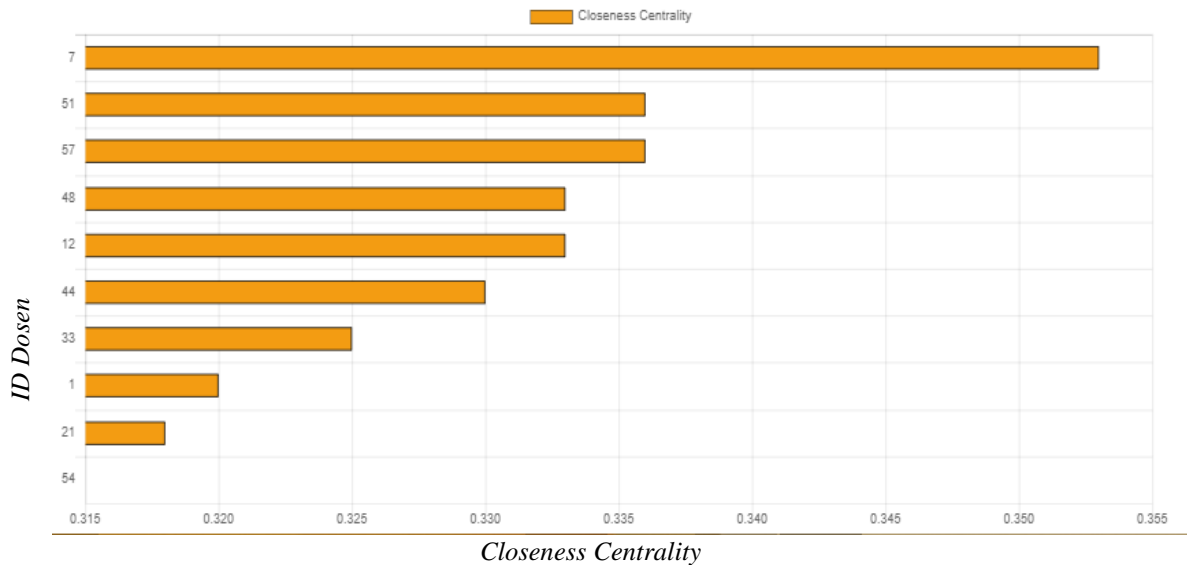


Gambar 2. Horizontal bar *degree centrality* dosen Politeknik Statistika STIS



Gambar 3. Horizontal bar *betweenness centrality* dosen Politeknik Statistika STIS

### Horizontal Bar Closeness Centrality Dosen Polstat Stis



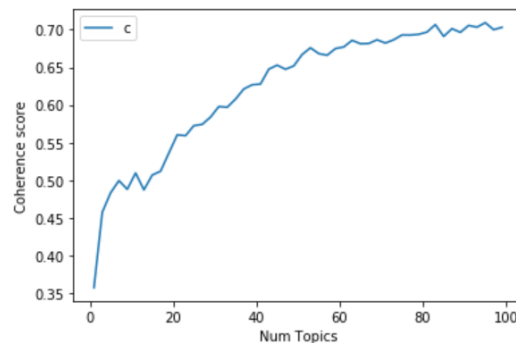
Gambar 4. Horizontal bar *closeness centrality* dosen Politeknik Statistika STIS

51 memiliki jumlah publikasi antara 21-25, dan memiliki jumlah koneksi/kolaborasi dengan dosen lain di lingkungan PolStat STIS paling banyak. Sedangkan dosen dengan jumlah publikasi terbanyak adalah dosen ID 48. Dosen ini terlihat tidak terlalu banyak memiliki jaringan/kolaborasi dengan internal dosen, kemungkinan dosen tersebut berkolaborasi dengan pihak luar Polstat STIS. Dosen ID 21 berkolaborasi cukup banyak dengan dosen lainnya, namun masih belum banyak memiliki jumlah publikasi.

Gambar 2, 3, dan 4 merupakan hasil dari metode evaluasi SNA menggunakan *package Python*. Gambar 2 menunjukkan bahwa dosen dengan kolaborasi terbanyak adalah dosen dengan ID 21 dan 51, yang kemudian terbanyak kedua adalah dosen ID 7. Sedangkan *betweenness centrality* yang ditampilkan di gambar 4 menunjukkan bahwa dosen yang menjadi penghubung terbanyak adalah dosen ID 44 diikuti oleh dosen ID 48. Berikutnya gambar 5 menunjukkan bahwa dosen yang menjadi pusat dari grafik adalah dosen dengan ID 7.

### Implementasi LDA

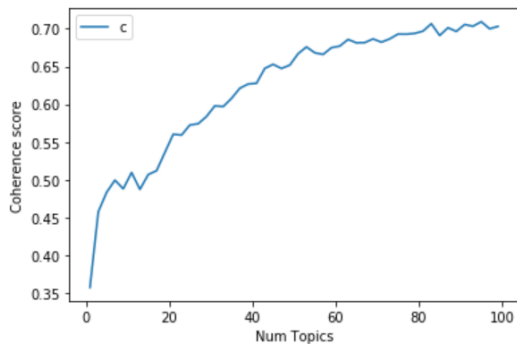
Evaluasi LDA dilakukan dengan cara mengecek nilai *perplexity* dan *coherence* dari model untuk menentukan jumlah topik yang sesuai.



Gambar 5. Uji *coherence* model LDA

Gambar 5 adalah hasil uji *coherence* dengan menggunakan modul *coherenceModel* dari *package gensim*. Metode *coherence* yang digunakan adalah *coherence* vektor. Berdasarkan gambar, terjadi kenaikan yang signifikan jumlah titik 1 ke titik 11 yaitu sebesar 42,6%, dari *coherence* 0,3574 menjadi 0,5096. Meski dari titik 11 nilai *coherence* terus mengalami kenaikan, agar terjadi kenaikan yang signifikan harus meningkatkan jumlah topik yang secara signifikan juga. Sehingga jumlah topik minimal dengan *coherence* terbaik adalah 11. *Coherence* menunjukkan nilai interpretabilitas, semakin tinggi maka akan semakin gampang dipahami manusia topik, sehingga *coherence* dengan nilai 0,5 masih dikatakan belum baik.

Gambar 6 merupakan grafik uji *perplexity* menggunakan fungsi *log perplexity* dari modul *LdaModel*. Semakin



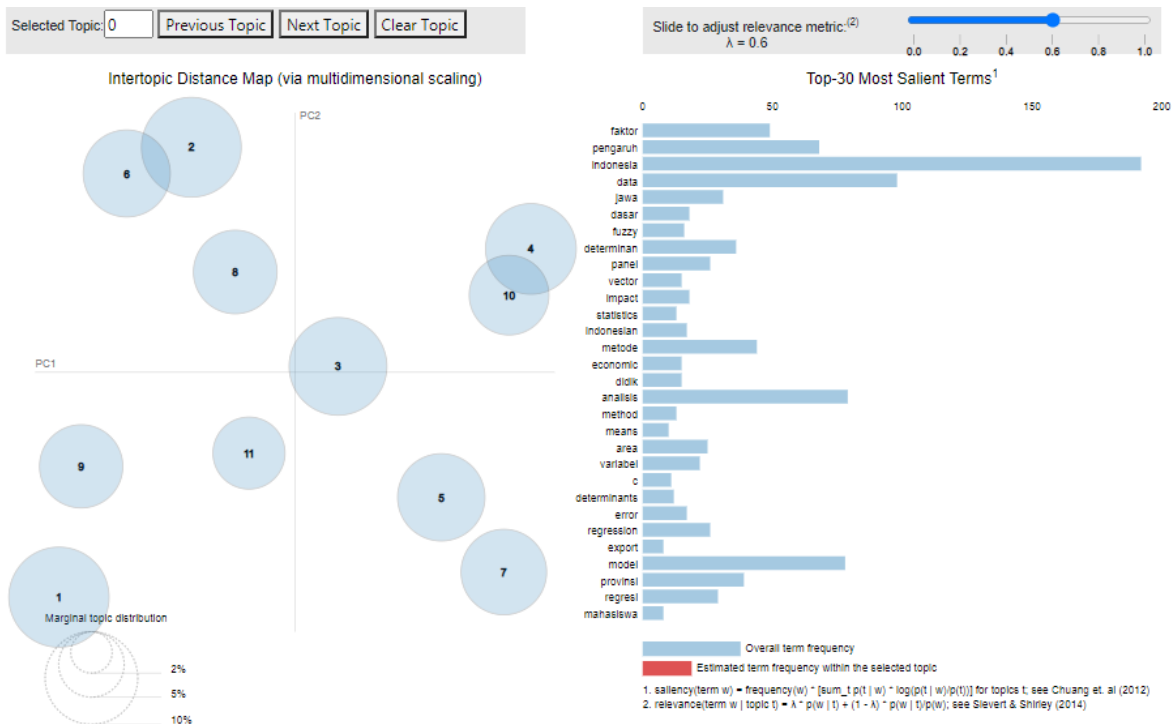
Gambar 6. Uji *perplexity* LDA model

kecil nilai *perplexity* menunjukkan bahwa model semakin baik dalam menangkap data. Dari gambar di atas terjadi penurunan signifikan hingga topik 17 dengan nilai *perplexity* sebesar -6.8179. Jumlah topik 15 dengan nilai *perplexity* sebesar -6.717 kurang bagus digunakan karena nilainya naik dan jumlah topik 19 dengan nilai *perplexity* sebesar -6.8173 juga kurang bagus karena nilainya tidak berbeda dengan jumlah topik 17. Tetapi dengan acuan uji *coherence* maka jumlah topik 11 dengan *perplexity* sebesar -6.6784 masih untuk digunakan.

Gambar 7 menampilkan grafik LDA judul publikasi dosen Politeknik Statistika STIS. Panel kiri membahas tentang topik yang dihasilkan. Terlihat pada gambar

terdapat 11 lingkaran yang menggambarkan 11 topik umum judul publikasi. Jarak antar lingkaran menggambarkan hubungan antar topik, semakin dekat jaraknya maka topiknya memiliki kesamaan. Ukuran lingkaran menunjukkan prevalensi topik atau ukuran dari topik, semakin besar ukuran lingkaran maka semakin banyaknya *token* yang dicakupnya, urutan topik sendiri diurutkan berdasarkan prevalensi tersebut. Tiga topik utama yang didapatkan dari penelitian dosen Politeknik Statistika STIS adalah faktor-faktor tertentu yang memengaruhi variabel tertentu dalam analisis regresi, pembahasan mengenai data rumah tangga, pangan, dan kemiskinan di daerah Jawa Timur dan Nusa Tenggara Timur, dan penggunaan data panel untuk mengestimasi ekonomi di Indonesia.

Kontribusi penelitian ini berupa penerapan metode SNA dan LDA menggunakan *python* untuk mengetahui perkembangan penelitian yang dilakukan oleh dosen Politeknik Statistika STIS, baik dari topik maupun kolaborasi penelitian mereka. Untuk pendekatan dengan SNA merupakan penyederhanaan kolaborasi nama penulis sehingga hanya nama dosen dari Instansi tertentu yang tercakup dan



Gambar 7. Grafik LDA judul publikasi dosen Politeknik Statistika STIS



penambahan informasi tertentu pada grafik jaring seperti jumlah publikasi yang dilakukan. Sedangkan dalam implementasi LDA, terdapat keterbatasan dikarenakan kekurangan data untuk diolah menyebabkan pendefinisian topik tidak dapat dilakukan karena topik yang dihasilkan tidak membentuk sebuah kalimat atau topik.

## KESIMPULAN

Berdasarkan pembahasan yang telah disampaikan sebelumnya, maka peneliti menyimpulkan sebagai berikut:

1. Pengimplementasian SNA terhadap data publikasi ilmiah dosen Politeknik Statistika STIS berhasil menghasilkan dan menampilkan kolaborasi penelitian. Dengan menggunakan pengukuran sentralitas dapat diidentifikasi dosen-dosen yang berposisi sebagai sentral dalam kolaborasi penelitian antar dosen.
2. Pengimplementasian LDA terhadap publikasi ilmiah dosen Politeknik Statistika STIS telah menampilkan beberapa topik utama pada publikasi ilmiah dosen berdasarkan distribusi kata pada judul publikasi.

## SARAN

Disebabkan keterbatasan penelitian ini, peneliti mengusulkan beberapa saran sebagai berikut untuk pengembangan berikutnya:

1. Melakukan pengecekan manual yang lebih mendalam terhadap nama penulis untuk menemukan nama dosen yang salah di dalam nama penulis, dengan menggunakan metode *replace* nama yang salah tersebut bisa digantikan dengan nama yang sesuai standar yang digunakan.
2. Melakukan penerjemahan terhadap judul yang berbahasa Inggris ke dalam Bahasa Indonesia agar tidak terdapat topik yang terbentuk karena kesamaan bahasa yang digunakan.
3. Melakukan penurunan jumlah topik jika model yang digunakan masih sama, karena berdasarkan hasil penelitian ini penggunaan jumlah topik 11 menghasilkan beberapa topik yang tidak bagus.

4. Menggunakan data abstrak dan kata kunci dari publikasi ilmiah dosen untuk diterapkan metode LDA agar datanya lebih banyak dan variatif

## DAFTAR PUSTAKA

- Avasarala, S. 2014. *Selenium Webdriver Practical Guide*. Birmingham: Packt Publishing Ltd.
- Bavelas, A. 1948. A Mathematical Model for Group Structures. Human Organization.
- Beauchamp, M.A. 1965. An improved index of centrality. Syst. Res.
- Blei, D. M, Ng, A.Y, & Jordan, M.I. 2003. *Latent Dirichlet Allocation*. Berkeley: University of California.
- Carrington, P., Scott, J., & S. Wasserman. (Eds.). 2005. *Models and Methods in Social Network Analysis* (Structural Analysis in the Social Sciences). Cambridge: Cambridge University Press.
- de Finetti, B. 1990. *Theory of Probability* (A critical introductory treatment). Wiley, New York.
- Dvorski, D. D. 2007. *Installing, Configuring, and Developing with Xampp*. Ontario: Skills Canada.
- Everett, M and Borgatti, S.P. 2003. *Extending Centrality*. London: University of Westminster.
- Feng, Jun, dkk. 2019. *Product Feature Extraction via Topic Model and Synonym Recognition Approach*. Nanjing: Hohai University.
- Freeman, L. C. 1979. *Centrality in Social Networks Conceptual Clarification*. Lehigh University.
- Hagberg, A. A. dkk. 2008. *Exploring Network Structure, Dynamics, and Function using NetworkX*. Los Alamos: Los Alamos National Laboratory.
- Hakimi, S. 1965. *Optimal Distribution of Switching Centers in a Communication Network and Some Related Theoretic Graph Theoretic Problems*. Evanston, Illinois
- Hunt, J. 2019. *Advanced Guide to Python 3 Programming*. Cham: Springer Nature Switzerland AG.



- Anthonisse, J.M. 1971. *The Rush in a Directed Graph*, diterbitkan Stichting Mathematisch Centrum. Mathematische Besliskunde. Amsterdam, Netherlands
- Leavitt, H. J. 1951. Some effects of certain communication patterns on group performance. *The Journal of Abnormal and Social Psychology*
- Loper, E & Bird, S. 2002. *NLTK: The Natural Language Toolkit*. Philadelphia: University of Pennsylvania.
- Putra, I. M. K. B. 2017. Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA). Surabaya: Institut Teknologi Sepuluh Nopember.
- Baeza-Yates, R., dan Ribeiro-Neto, B. 1999. *Modern Information Retrieval: The Concepts and Technology behind Search*
- Roder, M., Both, A., dan Hinneburg, A.. 2015. *Exploring the Space of Topic Coherence Measures*. Leipzig: Leipzig University.
- Sabidussi, G. 1996. The centrality index of a graph. *Psychometrika* **31**.
- Shaw, M. E. 1954. Some effects of unequal distribution of information upon group performance in various communication nets.
- Sievert, C & Shirley K.E. 2014. *LDAvis: A method for visualizing and interpreting topics*. Ames: Iowa State University.
- Suehring, S. 2002. *MySQL Bible*. New York: Wiley Publishing, Inc.
- Pryke, S.D. 2004. Analysing construction project coalitions: exploring the application of social network analysis, *Construction Management and Economics*, 22:8, 787-797, DOI:
- Tuhuteru, H., dan Iriane, A. 2018. Analisis Kolaborasi Penelitian Ilmiah Dosen Fakultas X dengan Social Network Analysis (SNA). Salatiga: Universitas Kristen Satya Wacana.
- Wasserman, S., dan Faust, K. 1994. *Structural analysis in the social sciences*. Social network analysis: Methods and applications. Cambridge University Press.
- Zhu, L., Menzies, N.A., Wang, J. *et al.* 2020. Estimation and correction of bias in network simulations based on respondent-driven sampling data. *Sci Rep* 10.



# INTERPOLASI CURAH HUJAN EKSTRIM MENGGUNAKAN MODEL SPATIAL DI PROVINSI JAWA BARAT

Rahmi Lathifah Islami<sup>1</sup>, Pardomuan Robinson Sihombing<sup>2</sup>

<sup>1</sup>Universitas Padjadjaran, <sup>2</sup>Badan Pusat Statistik  
e-mail: [2robin\\_sihombing@yahoo.com](mailto:2robin_sihombing@yahoo.com)

## Abstrak

Curah hujan ekstrim merupakan aspek yang penting untuk dikaji akibat dampaknya yang dapat merugikan. Kendala yang ada adalah belum tentu ada nilai ekstrim di setiap lokasi, sehingga diperlukan suatu pewilayahan yang dapat menjelaskan secara khusus mengenai curah hujan ekstrim dari segi dependensi lokasi, serta pola hujan ekstrim secara spasial. Tujuan penelitian, ini membangun suatu perwilayahan khusus curah hujan ekstrim di wilayah Jawa Barat pada periode Desember 2017-Februari 2018. Salah satu metode yang digunakan untuk menduga data curah hujan yang tidak lengkap adalah dengan teknik interpolasi. Terdapat beberapa metode yang bisa digunakan untuk melakukan interpolasi di antaranya Kriging dan Thin Plate Splines. Dalam penelitian ini akan membandingkan kedua metode tersebut berdasarkan nilai GVC-nya. Hasil yang didapat berdasarkan nilai GCV, metode Thin Plate Splines lebih baik digunakan untuk membangun perwilayahan khusus curah hujan ekstrim di wilayah Jawa Barat pada periode Desember 2017-Februari 2018.

**Kata kunci:** Hujan Ekstrim, Kriging, Thin Plate Splines

## Abstract

*Extreme rainfall is an important aspect to study due to its detrimental impacts. The existing constraint is that there is not necessarily an extreme value in every location, so that a zoning is needed that can specifically explain extreme rainfall in terms of location dependency, as well as spatial extreme rainfall patterns. The purpose of this research is to build a special area for extreme rainfall in the West Java region in the period December 2017-February 2018. One of the methods used to estimate incomplete rainfall data is the interpolation technique. There are several methods that can be used to perform interpolation, including Kriging and Thin Plate Splines. In this study, the two methods will be compared based on the GVC value. The results obtained are based on the GCV value, the Thin Plate Splines method is better used to build special areas for extreme rainfall in the West Java region in the period December 2017-February 2018.*

**Keywords:** Extreme Rain, Kriging, Thin Plate Splines

## PENDAHULUAN

Cuaca dan iklim adalah unsur geografis yang selalu menarik untuk dibahas karena sangat memengaruhi kehidupan manusia. Salah satu parameter cuaca dan iklim yang penting adalah curah hujan. Curah hujan memiliki peran penting misalkan saja dalam bidang pertanian, namun juga kerap menjadi sumber bencana. Tingginya curah hujan di wilayah Indonesia menyebabkan wilayah ini rentan terhadap bencana banjir. Jika dilihat dari dampak yang ditimbulkan maka kajian cuaca dan iklim ekstrim perlu dikembangkan di Indonesia (lintang rendah), khususnya Provinsi Jawa Barat karena dinamika atmosfer di wilayah Indonesia memang lebih sulit diprediksi dibandingkan negara-negara di lintang menengah dan tinggi.

Provinsi Jawa Barat memiliki kondisi topografi yang cukup kompleks sehingga pengaruhnya sangat besar terhadap variasi hujan secara spasial, adanya pegunungan yang berhadapan dengan sumber uap air seperti lautan akan meningkatkan curah hujan di wilayah pegunungan tersebut terutama pada bagian depan yang menghadap arah angin, karena pada wilayah tersebut uap air akan terangkat naik karena adanya gunung dan membentuk awan (Masruroh, 2013). Dengan demikian, penting untuk mempelajari informasi dan pengetahuan tentang cuaca dan iklim. Dengan mengetahui pola cuaca ekstrim maka dampak dari kejadian ekstrim tersebut dapat diantisipasi sedini mungkin. Selain itu pengetahuan yang bagus tentang cuaca dan iklim, terutama kejadian cuaca dan iklim juga berguna bagi bidang pertanian dan hidrologi.

Pada tahun 2017-2018 Badan Nasional Penanggulangan Bencana mencatat telah terjadi 152 kali kejadian banjir di Jawa Barat yang merupakan angka kasus tertinggi keempat diseluruh Indonesia. Hal tersebut terjadi karena curah hujan yang cukup tinggi di Jawa Barat. Berdasarkan amplitudo dan fase tahunan, Indonesia diklasifikasikan ke dalam empat wilayah klimatologi, yaitu: Desember-Januari-Februari (DJF), Maret-April-Mei

(MAM), Juni-Juli- Agustus (JJA) dan September-Oktober-November (SON). Jawa Barat merupakan wilayah klimatologi DJF (Hendon, 2003 dalam Rustiana, 2018). Hal tersebut juga didukung oleh penelitian Rinaldi (2018) yang diketahui curah hujan di Jawa barat memiliki fase ekstrim baik dalam hal frekuensi maupun curahnya selama bulan Desember-Januari-Februari.

Curah hujan ekstrim merupakan aspek yang penting untuk dikaji akibat dampaknya merugikan. Jika curah hujan yang ekstrim terjadi maka salah satu dampaknya adalah bencana banjir dan tanah longsor yang dapat merugikan baik bagi masyarakat, sehingga penting untuk diketahui tentang kondisi cuaca hujan ekstrim. Kendala yang ada adalah belum tentu ada nilai ekstrim di setiap lokasi, sehingga diperlukan suatu pewilayahan yang dapat menjelaskan secara khusus mengenai curah hujan ekstrim dari segi dependensi lokasi, serta pola hujan ekstrim secara spasial. Permasalahan ini menjadi landasan tujuan penelitian, yaitu membangun suatu perwilayahan khusus curah hujan ekstrim di wilayah Jawa Barat.

Salah satu metode yang digunakan untuk menduga data curah hujan yang tidak lengkap adalah interpolasi. Interpolasi merupakan suatu metode untuk menghasilkan data baru dari data-data yang diketahui. Metode tersebut dapat digunakan dalam pendugaan curah hujan ekstrim karena setiap titik yang diinterpolasi dipengaruhi oleh nilai dari titik di sekitarnya (Pasaribu dan Haryani 2012).

Terdapat beberapa metode yang bisa digunakan untuk melakukan interpolasi seperti *Kriging* dan *Thin Plate Splines*. Masing-masing metode ini akan memberikan hasil interpolasi yang berbeda. Akan menjadi mudah dan bermanfaat bagi pengguna berikutnya misalnya peneliti selanjutnya apabila ada kajian tentang perbandingan hasil interpolasi dengan metode yang berbeda sehingga metode yang tepat bisa dipilih. Perwilayahan curah hujan dari hasil interpolasi dapat digunakan untuk menggambarkan pola curah hujan (Militano, 2015). Penelitian yang mengkaji tentang pembentukan area untuk curah

hujan ekstrim merupakan hal yang sangat penting. Oleh sebab itu, tujuan dalam penelitian ini adalah mengkaji pembentukan wilayah berdasarkan curah hujan ekstrim dengan menggunakan metode *Kriging* dan *Thin Plate Splines*.

## METODOLOGI

### Data dan Variabel Penelitian

Data yang digunakan adalah data sekunder yang diperoleh dari Badan Meteorologi Klimatologi dan Geofisika (BMKG). Data ini berupa data jumlah curah hujan bulanan, koordinat stasiun dan ketinggian di atas permukaan laut. Jumlah stasiun/pos curah hujan bulanan yang digunakan 17 stasiun/pos stasiun hujan di Jawa Barat. Periode data curah hujan yang digunakan adalah dari Desember 2017 sampai tahun Februari 2018.

### Curah Hujan Ekstrim

Curah hujan ekstrim adalah kondisi curah hujan yang cukup tinggi/rendah dari rata-rata kondisi normalnya. Menurut Supriatna dalam BMKG (2011) curah hujan dengan intensitas > 100 mm/hari menjadi parameter terjadinya hujan ekstrim. Jawa barat memiliki fase ekstrim di bulan Desember-Januari-Februari (Rinaldi, 2018).

### Variogram

Variogram merupakan plot dua dimensi yang digunakan untuk mengukur keragaman spasial antar amatan pada data yang digunakan. Variogram mencerminkan karakteristik dari korelasi spasial, yaitu korelasi antara dua buah amatan menjadi berkurang atau tidak berkorelasi karena bertambahnya jarak antar amatan tersebut (Cressie, 1993). Terdapat dua metode estimator dalam *package geoR*.

a. Classical estimator

$$\gamma(h) = \frac{1}{2N_h} \sum_{i=1}^{N_h} (Y(x_i + h) - Y(x_i))^2 \quad (1)$$

b. Hawkins and Crassie's modulus Estimator

$$\gamma(h) = \frac{\left( \frac{1}{N_h} \sum_{i=1}^{N_h} (Y(x_i+h) - Y(x_i))^2 \right)^4}{0.914 + \frac{0.988}{N_h}} \quad (2)$$

dengan:

- $x_i$  : vektor dari koordinat spasial pada titik ke-i
- $Y(x_i)$  : variabel pada titik ke-i
- $H$  : vektor lag antara dua amatan (lokasi)
- $Y(x_i + h)$  : nilai variabel setelah dilag kan pada titik ke-i
- $N_h$  : jumlah pasangan pengamatan pada lag h

### Kriging

Kriging merupakan interpolasi suatu nilai peubah pada suatu titik (lokasi) tertentu yang dilakukan dengan mengamati data yang sejenis di lokasi lainnya. Nilai dugaan diperoleh dengan rumus (Cressie 1993) yaitu:

$$z_i = \beta_0 1 + \beta_1 x_i + \beta_2 y_i + \beta_3 h_i + \beta_4 a_i + \epsilon_i \quad (3)$$

dengan:

- $\beta_0$ : Intersep
- $\beta_k$ : Intersep  $x_s$  dan  $y_s$ : Korordinat lokasi ke-i
- $h_i$ : Elevasi lokasi ke-i
- $a_i$ : Variabel pengamatan ke-i
- $\epsilon_i$  : error lokasi ke-i dengan  $\epsilon_i \sim normal(0, \sigma_i^2)$

### Thin Plate Spline

Thin plate spline merupakan metode spline untuk dimensi ganda, yang juga dikenal sebagai spline-d, dengan d menyatakan dimensi. Thin plate spline merupakan hasil dari meminimalkan residual sum of squares yang bergantung pada tingkat kehalusan (Duchon, 1977 dalam Militano 2015).

$$\frac{1}{n} \sum_{i=1}^n \left( z_{s_i} - f(x_i, y_i) \right)^2 + \lambda \iint \left[ \left( \frac{\partial^2 f}{\partial x^2} \right)^2 + \left( \frac{\partial^2 f}{\partial y^2} \right)^2 + 2 \left( \frac{\partial^2 f}{\partial x^2 \partial y^2} \right)^2 \right] dx dy \quad (4)$$

Parameter smoothing  $\lambda$  harus dipilih dengan tepat untuk memaksimalkan hasil. Thin plate spline yang digunakan pada penelitian ini mirip dengan model kriging, hanya saja hubungan curah hujan tidak digambarkan secara linier namun dengan hubungan berdasarkan fungsi kehalusan dengan mengasumsikan  $\epsilon_i \sim normal(0, \sigma_i^2)$ . Berikut adalah model dari thin plate spline

$$z_i = f(x_i, y_i) + \beta_1 h_i + \beta_2 a_i + \epsilon_i \quad (5)$$

### Generalized Cross Validation (GCV)

GCV merupakan suatu metode yang digunakan untuk menentukan bandwidth optimum dalam suatu model dengan fungsi sebagai berikut:

$$GCV(h) = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\left( \frac{1}{n} \text{tr} [I - (X^T K_h X)^{-1} X^T K_h] \right)^2} \quad (6)$$

### Langkah Penelitian

1. Eksplorasi data untuk menjelaskan statistik deskriptif dari data tanpa mempertimbangkan efek stasiun sehingga dapat mengetahui secara garis besar curah hujan ekstrim di Pulau Jawa.
2. Analisis variogram untuk melihat efek spasial dari data.
3. Analisis interpolasi model kriging dan thin plate splines.
4. Memilih model terbaik dengan melihat nilai GCV Sigma dan menginterpretasikannya

## HASIL DAN PEMBAHASAN

### Statistika Deskriptif

Curah hujan ekstrim adalah kondisi curah hujan yang cukup tinggi/rendah dari rata-rata kondisi normalnya. Curah hujan di Jawa barat memiliki fase ekstrim di bulan Desember-Januari-Februari yang dapat dilihat dari data memiliki nilai yang sangat tinggi.

Tabel 1. Statistik Deskriptif

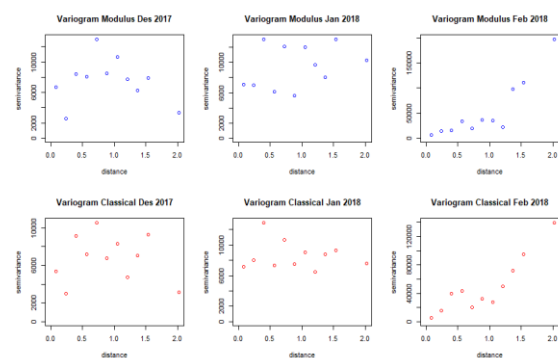
	Desember 2017	Januari 2018	Februari 2018
Min.	72	6	235
Median	206	147	364
Mean	210.6	160.5	444.1
Max.	350	373	904
Standar Deviasi	85.97	92.42	200.26

Berdasarkan Tabel 1 rata-rata curah hujan di Jawa Barat pada bulan Desember 2017 sebesar 210.6mm. Curah Hujan pada bulan Desember 2017 mencapai nilai tertinggi sebesar 350 mm yang tercatat di Stasiun Banjaran dan mengalami nilai terendah sebesar 72 mm yang tercatat di

Stasiun Ciawi. Sedangkan pada bulan Januari 2018 memiliki penurunan rata-rata yaitu sebesar 160.5 mm yang memiliki curah hujan minimum di Stasiun Cikeding sebesar 6 mm dan maksimum di Stasiun Cinangka sebesar 373mm. Berdasarkan data yang didapatkan pada ketiga periode waktu, bulan Februari 2018 mencapai nilai paling tinggi dengan rata-rata curah hujan di Jawa Barat sebesar 444.1 mm dengan nilai minimum diatas rata-rata bulan Desember 2017 yaitu 235 mm dan nilai maksimum sebesar 904 mm. Curah hujan terendah di Jawa Barat pada bulan Februari tercatat di Stasiun Cikeding dan curah hujan tertinggi tercatat di stasiun Kawali. Namun, berdasarkan sebaran data, bulan Februari memiliki sebaran data yang paling tinggi yang artinya nilai curah hujan di Jawa Barat berdasarkan 17 stasiun memiliki nilai yang cukup fluktuatif.

### Variogram

Setelah dilakukan eksplorasi data, selanjutnya ingin dilihat efek spasial stasiun curah hujan di Jawa Barat dengan tiga periode waktu, yaitu Desember 2017, Januari 2018, dan Februari 2018. Berikut hasil dari plot variogram yang telah diolah *software R*.



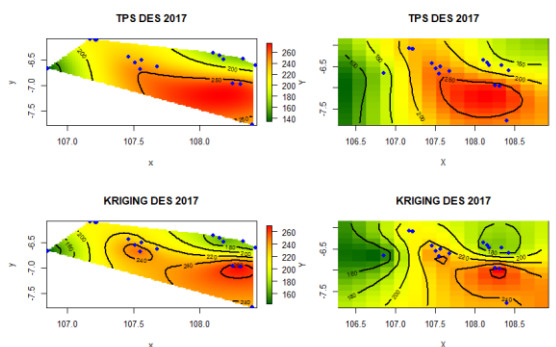
Gambar 1. Variogram Curah Hujan di Jawa Barat

Pada Gambar 1 telah diplot variogram dengan *classical Estimator* yang gambarkan dengan warna biru dan *Modulus Estimator* yang berwarna merah. Berdasarkan hasil plot dapat dilihat pada bulan Desember 2017 sebaran data berdasarkan jarak terlihat membentuk kurva melengkung ke bawah sementara pada bulan Januari 2018 bersifat konstan dan

pada bulan Februari 2018 meningkat. Hal tersebut mengakibatkan data pada bulan Januari tidak dapat dilakukan analisis lebih lanjut karena jarak tidak mempengaruhi curah hujan yang berarti.

### Interpolasi Curah Hujan

Setelah dilakukan analisa menggunakan variogram selanjutnya melakukan interpolasi menggunakan metode *kriging* dan *thin plate spline*. Berdasarkan hasil dari *software* R adalah sebagai berikut:

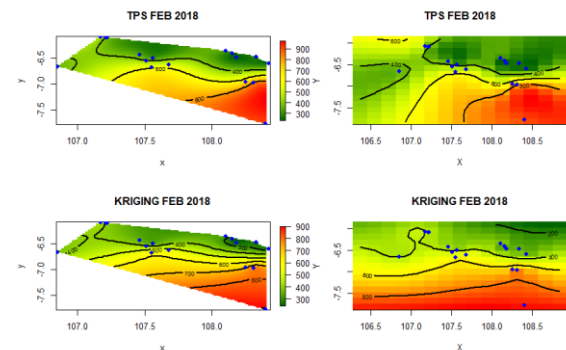


Gambar 2. Interpolasi Curah Hujan di Jawa Barat Desember 2017

Pada Gambar 2, interpolasi menggunakan *thin plate spline* berada disisi atas dan *kriging* berada dibawah. Gambar sisi kiri merupakan hasil otomatis yang dikeluarkan oleh *software* R ketika menggunakan *function* **Tps** dan **Krig**. Berdasarkan gambar terlihat hasil interpolasi belum mencakup seluruh wilayah yang terdapat di Jawa Barat sehingga intepolasi pun diperlebar berdasarkan titik koordinat maksimum dan minimum dari Provinsi Jawa Barat yang terlihat pada gambar disisi sebelah kanan. Berdasarkan hasil interpolasi curah hujan di Jawa Barat pada bulan Desember 2017 yang terlihat disisi kanan dapat dilihat terlihat kemiripan perwilayahan dari hasil *thin plate spline* dan *kriging*. Namun, hasil perwilayahan dari *thin plate spline* terlihat sebagian daerah Jawa Barat memiliki nilai curah hujan yang cukup ekstrim jika dibandingkan dengan hasil perwilayahan dengan menggunakan *kriging*.

Berdasarkan pola curah hujan di bulan Desember 2017 pada Gambar 2, dengan melihat titik koordinat dapat

diidentifikasi Kabupaten Majalengka dan sekitarnya memiliki curah hujan yang tinggi yang ditandai dengan warna merah. Curah hujan dengan intensitas tinggi juga sampai ke Kabupaten Karawang dan Bandung. Sementara Kabupaten Indramayu mengalai curah hujan yang cukup sedang ditandai dengan pola gradasi warna hijau. Sedangkan warna kuning mendominasi di wilayah Cianjur Bogor dan Bekasi yang memiliki curah hujan yang cukup tinggi.



Gambar 3. Interpolasi Curah Hujan di Jawa Barat Februari 2018

Hasil interpolasi curah hujan di Jawa Barat pada bulan Februari 2018 terlihat perbedaan yang cukup signifikan antara metode *thin plate spline* dan *kriging*. Pada perwilayahan yang dihasilkan oleh metode *thin plate spline* dan *kriging* pada Gambar 3 awalnya terlihat mirip seperti yang terlihat di gambar sisi kanan, namun ketika jarak perwilayahan diperlebar untuk menginterpolasi secara lengkap curah hujan di Jawa Barat terlihat bahwasannya pada metode *kriging* semakin rendah nilai koordinat Bujur dan semakin tinggi nilai curah hujan di Jawa Barat. Hal tersebut tentu berbeda dengan hasil perwilayahn yang didapatkan menggunakan metode *thin plate spline* yang memiliki pola tidak linier seperti yang dihasilkan oleh metode *kriging*.

Berdasarkan Gambar 3 pola curah hujan pada bulan Februari 2018 di Kabupaten Ciamis, Majalengka dan sekitarnya mengalami curah hujan yang paling tinggi ditandai dengan warna merah yang berada pada Gambar 3. Kabupaten Purwakarta, Karawang dan Subang memiliki intensitas curah hujan yang berkisar 600mm. Sedangkan Kabupaten

Indramayu, Cirebon dan sekitarnya memiliki curah hujan yang related rendah jika dibandingkan kabupaten lainnya yang ditandai dengan warna hijau pada Gambar 3.

Setelah dilakukan interpolasi curah hujan di Jawa Barat pada bulan Desember 2017 dan Februari 2018 tentu ingin diketahui mana metode yang paling sesuai dengan data curah hujan di Jawa Barat. Berdasarkan hasil olahan *software* R didapatkan hasil nilai GCV dari sigma (*nugget standard deviation*) sebagai berikut:

Tabel 2. Nilai CGV

	<i>Kriging</i>	<i>Thin plate spline</i>
<i>Desember 2017</i>	73.12	75.07
<i>Februari 2018</i>	36.12	38.03

Berdasarkan tabel tersebut dapat dilihat bahwasannya dari kedua periode yaitu Desember 2017 dan Februari 2018 hasil interpolasi yang paling sesuai terhadap data curah hujan di Jawa Barat adalah dengan menggunakan metode *thin plate spline* karena memiliki nilai kesesuaian perwilayahan yang lebih tinggi.

## KESIMPULAN

Provinsi Jawa Barat memiliki kondisi topografi yang cukup kompleks sehingga pengaruhnya sangat besar terhadap variasi hujan ekstrim secara spasial. Kendala yang ada adalah belum tentu ada nilai ekstrim di setiap lokasi sehingga dilakukan interpolasi untuk melihat perwilayahan curah hujan di Jawa Barat. Periode waktu yang digunakan untuk menganalisa curah hujan ekstrim yaitu bulan Desember 2017, Januari 2018, dan Februari 2018. Metode interpolasi yang digunakan adalah *thin plate spline* dan *kriging*. Setelah dilakukan plot variogram data curah hujan Januari 2018 tidak bisa dianalisis lebih lanjut, sehingga yang dilakukan analisa interpolasi adalah curah hujan pada bulan Desember 2017 dan Februari 2018. Berdasarkan hasil interpolasi metode *thin plate spline* paling

sesuai dengan data curah hujan di Jawa Barat karena memiliki nilai kesesuaian perwilayahan yang lebih tinggi.

## DAFTAR PUSTAKA

- Badan Meteorologi Klimatologi dan Geofisika (BMKG). 2018. *Buletin Informasi Iklim Jawa Barat*. Bogor: Stasiun Klimatologi Bogor.
- Coles, SG. 2001. *An Introduction to Statistical Modeling of Extreme Values*. London: Springer Verlag. <https://doi.org/10.1007/978-1-44713675-0>.
- Cressie, NAC.1993. *Statistics for Spatial Data*. New York: Wiley.
- Hanifah, A dan Endarwin. 2011. Analisis Intensitas Curah Hujan Wilayah Bandung pada Awal 2010, *Jurnal Meteorologi dan Geofisika* 12 (2): 145-149.
- Masruroh, L. 2013. Analisis Curah Hujan Harian Maksimum dan Ekstrim di Kabupaten Bogor [skripsi]. Bogor: Institut Pertanian Bogor.
- Militino, *et al.* 2015. Interpolation of daily rainfall using spatiotemporal models and clustering. *International Journal of Climatology*. 35: 1453–1464.
- Pasaribu, JM, Haryani, NS. 2012. Perbandingan Teknik Interpolasi DEM SRTM dengan Metode Inverse Distance Weighted (IDW), Natural Neighbor, dan Spline. *J Penginderaan Jauh*. 9(2): 126-139.
- Rinaldi, A. 2018. Pengembangan Model Spatio-Temporal Conditional Autoregressive untuk Pendugaan Curah Hujan Ekstrim di Wilayah Jawa Barat[tesis]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.
- Rustiana, S. 2018. Model Generalized Space Time Autoregressive with Arch Error (Gstar-Arch) dan Penerapannya Terhadap Data Curah Hujan Pada Wilayah Rawan Bencana di Pulau Jawa [tesis]. Bandung: Universitas Padjadjaran.
- Suciantini. 2004. Evaluasi Prakiraan Sifat Hujan dan Penyusunan Model Prediksi Musim: Studi Kasus



Kabupaten Indramayu [tesis]. Bogor:  
Program Pascasarjana, Institut  
Pertanian Bogor



# ANALISIS JUMLAH PENDUDUK MISKIN DI PROVINSI MALUKU DENGAN MENGGUNAKAN PENDEKATAN REGRESI SPASIAL

Salmon Notje Aulele<sup>1</sup>, V. Y. I. Ilwaru<sup>2</sup>, E. R. Wuritimur<sup>3</sup>, M. Y. Matdoan<sup>4</sup>

<sup>1,2,3,4</sup>Universitas Pattimura  
e-mail: <sup>1</sup>keepyahya@gmail.com

## Abstrak

Provinsi Maluku berada pada urutan ke-4 Provinsi termiskin di Indonesia dengan memiliki persentase tingkat kemiskinan sebesar 18,45% dari total penduduk di Maluku. Aspek spasial perlu dikaji dikarenakan karakteristik setiap daerah atau wilayah yang berbeda-beda sehingga informasi tentang karakteristik lokasi ini bisa ditangkap dengan menggunakan model regresi spasial. Tujuan dari penelitian ini yaitu mengkaji karakteristik kemiskinan di Provinsi Maluku, kemudian menentukan faktor-faktor yang berpengaruh terhadap kemiskinan di Provinsi Maluku dengan menggunakan regresi spasial, serta mendapatkan model terbaik untuk memodelkan kemiskinan di Provinsi Maluku. Hasil penelitian menunjukkan bahwa rata-rata persentase penduduk miskin di Provinsi Maluku yaitu sebesar 22,71%, dimana Kabupaten Maluku Barat Daya memiliki persentase tertinggi sebesar 31,01% dan Kota Ambon memiliki persentase terendah sebesar 4,64%. Sedangkan faktor-faktor yang signifikan mempengaruhi kemiskinan di Provinsi Maluku dengan menggunakan regresi spasial adalah persentase rumah tangga (RT) yang bahan bakar untuk memasaknya dari kayu (X1), persentase penduduk berumur 7 – 24 tahun yang tidak/belum pernah sekolah (X2), persentase penduduk berumur 7 – 24 tahun yang tidak sekolah lagi (X3), persentase tingkat pengangguran terbuka (TPT) (X4) dan persentase tingkat partisipasi angkatan kerja (X5) dan model terbaik yang diperoleh untuk memodelkan kemiskinan di Provinsi Maluku yaitu *Spatial Lag Model*.

**Kata kunci:** Kemiskinan, Regresi Spasial, *Spatial Lag Model*

## Abstract

Maluku Province is in the 4th poorest province in Indonesia with a poverty rate of 18.45% of the total population in Maluku. The spatial aspect needs to be studied because the characteristics of each area or region are different so that information about the characteristics of this location can be captured using a spatial regression model. The purpose of this study is to examine the characteristics of poverty in Maluku Province, then determine the factors that affect poverty in Maluku province using spatial regression, and get the best model to model poverty in Maluku Province. The results showed that the average percentage of poor people in Maluku province is 22.71%, where Southwest Maluku Regency has the highest percentage of 31.01% and Kota Ambon has the lowest percentage of 4.64%. While the significant factors affecting poverty in Maluku province by using spatial regression are the percentage of households (rt) whose cooking fuel from wood (X1), percentage of non-school-aged 7 - 24 years (X2) (X3), percentage of the unemployed (TP) (X4) and percentage of labor force participation rate (X5) and the best model obtained to model poverty in Maluku Province is *Spatial Lag Model*.

**Keywords:** Poverty, Spatial Regression, Spatial Weighted Matrix, Spatial Model Lag

## PENDAHULUAN

Kemiskinan merupakan suatu masalah besar di semua negara di Dunia. Suatu negara dikatakan maju jika mampu memberantas kemiskinan di negara tersebut. Indonesia sebagai negara berkembang dan memiliki jumlah penduduk yang besar tidak dapat terhindar dari masalah kemiskinan (BPS, 2002). Menurut BPS, angka kemiskinan di Indonesia mengalami penurunan dalam 7 tahun terakhir namun cenderung melambat. Masih banyak masyarakat miskin yang kekurangan pangan, hal tersebut dapat diketahui dengan kemampuan membeli masyarakat yang rendah, serta kurangnya ketersediaan pangan (BPS Provinsi Maluku, 2017)

Kemiskinan juga ditandai dengan terbatasnya akses layanan kesehatan masyarakat yang kurang diperhatikan, sehingga berdampak pada rendahnya daya tahan untuk mencari nafkah, dan terbatasnya kemampuan anak untuk tumbuh dan berkembang. Kemiskinan juga berpengaruh pada sektor pendidikan yang membuat masyarakat miskin tidak dapat bersekolah dengan baik (Leleury dan Tomasouw, 2019). Hal ini disebabkan oleh terbatasnya jumlah sarana dan prasarana serta mutu pendidikan. Selain itu, kurangnya kesempatan kerja dan keterbatasan modal serta ketrampilan yang dimiliki.

Berdasarkan hasil survei BPS tahun 2017, Provinsi Maluku berada pada urutan ke-4 Provinsi termiskin di Indonesia dengan persentase tingkat kemiskinan sebesar 18,45% dari total penduduk di Provinsi Maluku. Untuk itu, Pemerintah Provinsi Maluku harus mengambil kebijakan yang tepat sasaran untuk dapat menurunkan jumlah penduduk miskin. (BPS, Provinsi Maluku, 2017).

Penggunaan aspek spasial digunakan pada penelitian ini dikarenakan karakteristik setiap daerah atau wilayah yang berbeda-beda sehingga informasi tentang karakteristik lokasi ini bisa ditangkap dengan menggunakan analisis data spasial. Anselin (1988) juga telah

mengembangkan beberapa model spasial yang didasarkan atas teori ekonomi yakni spatial econometrics (Akaike, 1974). Model tersebut terdiri dari spatial lag, spatial error, model campuran antara spatial lag dan spatial error (Aulele, S.N, 2014). Komponen yang mendasar dari model spasial adalah matriks penimbang spasial, matriks inilah yang mencerminkan adanya hubungan antara satu region dengan region lainnya (Draper dan Smith, 2016) Banyak metode dalam membuat matriks penimbang (Anton Howard, 1992). Metode yang biasa digunakan adalah pendekatan titik dan pendekatan Area. Pendekatan titik yaitu letak geografis suatu wilayah yang berdasarkan posisi koordinat garis lintang dan garis bujur. Pendekatan area berupa contiguity murni, ataupun memasukan unsur lain ke dalam konsep contiguity sesuai dengan kasus yang dipelajari (Anselin L, 1992). Karakteristik daerah yang beragam dan berbeda beda satu sama lainnya, sehingga perlu diakomodir dalam pembuatan suatu model. Dengan menggunakan model regresi spasial diharapkan mampu menghasilkan model kemiskinan yang spesifik di setiap daerah sehingga semakin bermanfaat dalam menurunkan jumlah penduduk miskin di Provinsi Maluku.

## METODE PENELITIAN

### Tipe Penelitian

Penelitian ini menggunakan tipe penelitian kuantitatif dengan pendekatan regresi spasial untuk menganalisis persentase penduduk miskin di Provinsi Maluku.

### Bahan atau Materi Penelitian

Data yang digunakan dalam penelitian data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Maluku tahun 2017. Pada penelitian ini yang dijadikan unit observasi adalah Kabupaten/Kota di Provinsi Maluku yaitu Kota Ambon, Kota Tual, Kabupaten Maluku Tengah, Kabupaten Maluku Tenggara Barat (MTB), Kabupaten Maluku Tenggara, Kabupaten Seram Bagian Timur

(SBT), Kabupaten Buru, Kabupaten Seram Bagian Barat (SBB), Kabupaten Maluku Barat Daya (MBD) dan Kabupaten Buru Selatan. Adapun software yang digunakan dalam penelitian ini yaitu dengan menggunakan *software* R dan Arcview.

### Prosedur Penelitian

Langkah-langkah dalam analisis data untuk setiap tujuan penelitian yaitu :

1. Mengkaji karakteristik penduduk berdasarkan deskriptif statistik dari variabel - variabel yang mempengaruhi kemiskinan di Provinsi Maluku
2. Untuk mendapatkan model regresi spasial berdasarkan faktor - faktor yang mempengaruhinya di Provinsi Maluku maka dilakukan analisis dengan langkah-langkah sebagai berikut (Laswinia dan Chamid, 2016) :
  - a. Verifikasi Data : pada tahapan ini akan dilakukan verifikasi terhadap data yang telah ditetapkan
  - b. Uji Multikolinieritas : pada tahapan ini akan dilihat variabel prediktor yang layak digunakan untuk pembuatan model
  - c. Pemodelan Regresi Berganda
  - d. Pemodelan Regresi Spasial dengan tahapan sebagai berikut (Johnson dan Wichern, 2007) :
    1. Identifikasi awal model Spasial : pada tahapan ini akan dilakukan pengujian dependensi spasial dengan menggunakan metode Moran's I dan Lagrange Multiplier tes (LM-tes).
    2. Membuat Matriks Pembobot : pada tahapan ini akan ditetapkan 2 matriks pembobot spasial yaitu Rook Contiguity dan Queen Contiguity
    3. Membuat model spasial : pada tahapan ini dilakukan pemodelan regresi spasial lag kemudian dilakukan estimasi parameter model dengan menggunakan Maximum Likelihood Estiamtor, selanjutnya dilakukan pengujian signifikansi koefisien spasial.
  - e. Pemilihan model terbaik : pada tahapan ini akan dipilih model terbaik

dari antara model regresi berganda dengan spatial lag model dengan menggunakan kriteria R<sup>2</sup> dan Akaike Information Criterion (AIC)

- f. Pengujian asumsi residual dari model terbaik

## HASIL DAN PEMBAHASAN

### Karakteristik Penduduk di Provinsi Maluku

Provinsi Maluku merupakan salah satu provinsi di Indonesia yang berbentuk kepulauan. Provinsi Maluku terdiri dari 9 Kabupaten dan 2

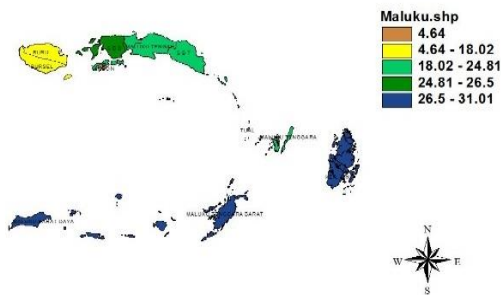
Kota dengan luas wilayah Provinsi Maluku secara keseluruhan adalah 712.480 km<sup>2</sup>. Dengan 92,4 persen wilayah Provinsi Maluku adalah lautan. Menurut letak geografis, wilayah Provinsi Maluku terletak antara 2<sup>o</sup> 30' - 9<sup>o</sup> Lintang Selatan dan 124<sup>o</sup> - 136<sup>o</sup> Bujur Timur. Selain sebagai provinsi kepulauan dengan potensi kelautan yang besar, Maluku sejak dahulu juga dikenal sebagai pulau penghasil rempah-rempah yang kaya.

Pada penelitian ini aplikasi model regresi spasial diterapkan pada kasus kemiskinan di Provinsi Maluku pada tahun 2016. Variabel yang diteliti yaitu persentase penduduk miskin tiap kabupaten/kota di provinsi Maluku sebagai variabel respon dan persentase rumah tangga (rt) yang bahan bakar untuk memasaknya dari kayu (X<sub>1</sub>), persentase penduduk berumur 7 – 24 tahun yang tidak/belum pernah sekolah (X<sub>2</sub>), persentase penduduk berumur 7 – 24 tahun yang tidak sekolah lagi (X<sub>3</sub>), persentase tingkat pengangguran terbuka (TPT) (X<sub>4</sub>) dan persentase tingkat partisipasi angkatan kerja (X<sub>5</sub>) tiap kabupaten/kota di Provinsi Maluku sebagai variabel prediktornya. Berikut kondisi masing-masing variabel di Provinsi Maluku.

#### a. Persentase Penduduk Miskin

Menurut BPS, penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan di bawah garis kemiskinan. Saat ini Provinsi Maluku berada pada peringkat ke-4 provinsi

termiskin di Indonesia. Persentase penduduk miskin tiap kabupaten/kota di provinsi Maluku disajikan dalam Gambar 1 dengan menggunakan *software* Arview sebagai berikut :



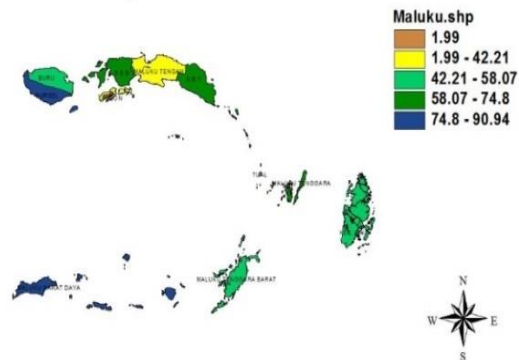
Gambar 1. Persentase Penduduk Miskin di Provinsi Maluku

Data penelitian menunjukkan bahwa rata-rata persentase penduduk miskin di Maluku yaitu 22,71%, dimana Kabupaten Maluku Barat Daya memiliki persentase tertinggi sebesar 31,01% dan Kota Ambon memiliki persentase terendah sebesar 4,64%. Berdasarkan Gambar 1, terlihat bahwa kabupaten/kota yang memiliki persentase penduduk miskin yang lebih dari 4,64% dan kurang dari sama dengan 18,02% adalah Kabupaten Buru dan Kabupaten Buru Selatan. Kabupaten/kota yang memiliki persentase penduduk miskin yang lebih dari 18,02% dan kurang dari sama dengan 24,81% adalah Kabupaten Maluku Tenggara, kabupaten Seram Bagian Timur, Kabupaten Maluku Tengah dan kota Tual. Kabupaten/kota yang memiliki persentase penduduk miskin yang lebih dari 24,81% dan kurang dari sama dengan 26,5% adalah Kabupaten Seram Bagian Barat. Kabupaten/Kota yang memiliki persentase penduduk miskin yang lebih dari 26,5% dan kurang dari sama dengan 31,01% adalah Kabupaten Maluku Barat Daya dan Kabupaten Maluku Tenggara Barat.

#### **b. Persentase Rumah Tangga yang Bahan Bakar Untuk Memasaknya dari Kayu**

Maluku menjadi salah satu Provinsi yang tertinggal, ditandai dengan masih ada keluarga atau masyarakat yang menggunakan kayu sebagai bahan bakar

utama untuk memasak sehari-hari, disaat bahan bakar minyak (BBM) telah disubsidi oleh pemerintah pusat. Berikut disajikan persentase rt yang bahan bakar untuk memasak dari kayu di Provinsi Maluku dalam Gambar 2 berikut :

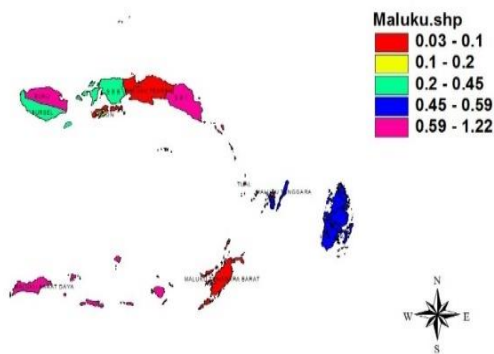


Gambar 2. Persentase Rumah Tangga yang bahan bakar untuk memasaknya dari kayu

Data penelitian menunjukkan bahwa rata-rata persentase RT yang bahan bakar untuk memasaknya dari kayu di Maluku yaitu 58,07%, dimana Kabupaten Maluku Barat Daya memiliki persentase tertinggi sebesar 90,94% dan Kota Ambon memiliki persentase terendah sebesar 1,99%. Berdasarkan Gambar 2, terlihat bahwa kabupaten/kota yang memiliki persentase RT yang bahan bakar untuk memasaknya dari kayu yang lebih dari 1,99% dan kurang dari sama dengan 42,21% adalah Kabupaten Maluku Tengah dan Kota Tual. Kabupaten/kota yang memiliki persentase RT yang bahan bakar untuk memasaknya dari kayu yang lebih dari 42,21% dan kurang dari sama dengan 58,07% adalah Kabupaten Buru dan Kabupaten Kepulauan Aru. Kabupaten/Kota yang memiliki persentase RT yang bahan bakar untuk memasaknya dari kayu yang lebih dari 58,07% dan kurang dari sama dengan 74,8% adalah kabupaten Seram Bagian Barat dan Seram Bagian Timur. Kabupaten/kota yang memiliki persentase RT yang bahan bakar untuk memasaknya dari kayu yang lebih dari 74,8% dan kurang dari sama dengan 90,94% adalah Kabupaten Buru Selatan.

### c. Persentase Penduduk Berumur 7 – 24 Tahun yang Tidak/Belum Pernah Sekolah

Realita yang terjadi, Maluku sebagai provinsi termiskin ke-4 di Indonesia, dapat dilihat juga dalam sektor pendidikan dengan presentase penduduk berumur 7 – 24 tahun yang tidak/belum pernah sekolah dalam Gambar 3 sebagai berikut:



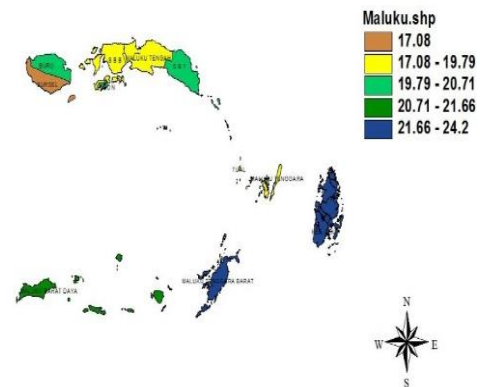
Gambar 3. Persentase penduduk berumur 7-24 tahun yang tidak/belum pernah sekolah

Data penelitian menunjukkan bahwa rata-rata persentase penduduk berumur 7-24 tahun yang tidak/belum pernah sekolah yaitu 0,49%, dimana Kabupaten Seram Bagian Timur memiliki persentase tertinggi sebesar 1,22% dan Kota Tual memiliki persentase terendah sebesar 0,03%. Berdasarkan Gambar 3, terlihat bahwa kabupaten/kota yang memiliki persentase penduduk berumur 7-24 tahun yang tidak/belum pernah sekolah yang lebih dari 0,03% dan kurang dari sama dengan 0,1% adalah Kabupaten Maluku Tengah dan kabupaten Maluku Tenggara Barat. Kabupaten/Kota yang memiliki persentase penduduk berumur 7-24 tahun yang tidak/belum pernah sekolah yang lebih dari 0,1% dan kurang dari sama dengan 0,2% adalah Kota Ambon. Kabupaten/Kota yang memiliki persentase penduduk berumur 7-24 tahun yang tidak/belum pernah sekolah yang lebih dari 0,2% dan kurang dari sama dengan 0,45% adalah Kabupaten Seram Bagian Barat dan Kabupaten Buru Selatan. Kabupaten/Kota yang memiliki persentase penduduk berumur 7-24 tahun yang tidak/belum pernah sekolah yang lebih dari 0,45% dan kurang dari sama dengan 0,61%

adalah Kabupaten Maluku Tenggara dan Kabupaten Kepulauan Aru. Kabupaten/Kota yang memiliki persentase penduduk berumur 7-24 tahun yang tidak/belum pernah sekolah yang lebih dari 0,61% dan kurang dari sama dengan 1,22% adalah Kabupaten Maluku Tenggara Barat.

### d. Persentase Penduduk Berumur 7–24 Tahun yang Tidak Sekolah Lagi

Salah satu faktor yang mempengaruhi kemiskinan pada sektor pendidikan adalah masih banyaknya anak usia sekolah yang berhenti sekolah. Berikut adalah persentase penduduk berumur 7–24 tahun yang tidak sekolah lagi pada Gambar 4, sebagai berikut:



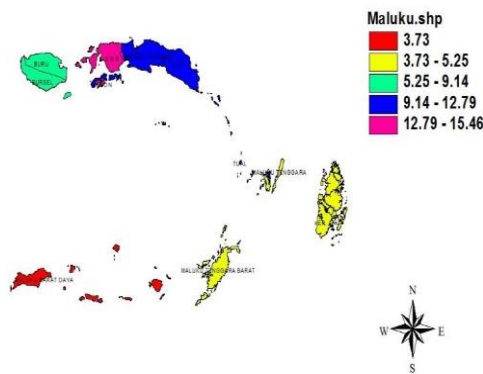
Gambar 4. Persentase penduduk berumur 7-24 tahun yang tidak sekolah lagi

Data penelitian menunjukkan bahwa rata-rata persentase penduduk berumur 7-24 tahun yang tidak sekolah lagi yaitu 20,50%, dimana Kabupaten Maluku Tenggara Barat memiliki persentase tertinggi sebesar 24,2% dan Kabupaten Buru Selatan memiliki persentase terendah sebesar 17,08%. Berdasarkan Gambar 4, terlihat bahwa kabupaten/kota yang memiliki persentase penduduk berumur 7-24 tahun yang tidak sekolah lagi yang lebih dari 17,08% dan kurang dari sama dengan 19,79% adalah Kabupaten Maluku Tenggara, Kabupaten Maluku Tengah, Kabupaten Seram Bagian Barat dan Kota Tual. Kabupaten/kota yang memiliki persentase penduduk berumur 7-24 tahun yang tidak sekolah lagi yang lebih dari 19,79% dan kurang dari sama dengan 20,71% adalah Kabupaten Buru, Kabupaten Seram Bagian Timur, Kabupaten Maluku

Tengah dan Kota Ambon. Kabupaten/kota yang memiliki persentase penduduk berumur 7-24 tahun yang tidak sekolah lagi yang lebih dari 20,71% dan kurang dari sama dengan 21,66% adalah Kabupaten Maluku Barat Daya. Kabupaten/kota yang memiliki persentase penduduk berumur 7-24 tahun yang tidak sekolah lagi yang lebih dari 21,66% dan kurang dari sama dengan 24,2% adalah Kepulauan Aru.

**e. Persentase Tingkat Pengangguran Terbuka (TPT)**

TPT yang tinggi menunjukkan bahwa terdapat banyak angkatan kerja yang tidak terserap pada pasar. Berikut adalah presentase tingkat pengangguran terbuka di Maluku disajikan dalam Gambar 5, sebagai berikut:



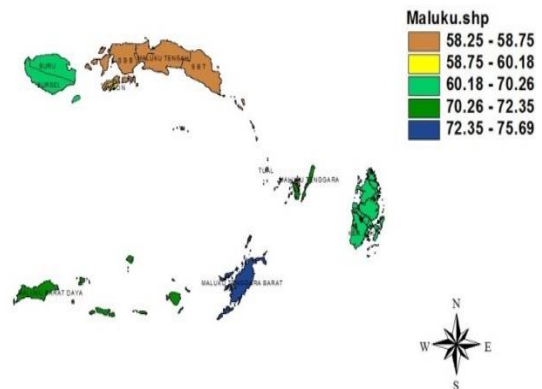
Gambar 5. Persentase Tingkat Pengangguran Terbuka (TPT)

Data penelitian menunjukkan bahwa rata-rata persentase tingkat pengangguran terbuka (TPT) di Provinsi Maluku yaitu 9,31%, dimana kota Ambon memiliki persentase tertinggi sebesar 15,46% dan Kabupaten Maluku Barat Daya memiliki persentase terendah sebesar 3,73%. Berdasarkan Gambar 5, terlihat bahwa kabupaten/kota yang memiliki persentase TPT yang lebih dari 3,73% dan kurang dari sama dengan 5,25% adalah Kabupaten Maluku Tenggara, Kabupaten Maluku Tenggara dan Kabupaten Kepulauan Aru. Kabupaten/kota yang memiliki persentase TPT yang lebih dari 5,25% dan kurang dari sama dengan 9,14% adalah Kabupaten Buru dan kabupaten Buru Selatan. Kabupaten/kota yang memiliki persentase TPT yang lebih dari 9,14% dan kurang dari

sama dengan 12,79% adalah Kabupaten Seram Bagian Timur, Kabupaten Maluku Tengah dan Kota Tual. Kabupaten/kota yang memiliki persentase penduduk miskin yang lebih dari 12,79% dan kurang dari sama dengan 15,46% adalah Kabupaten Seram Bagian Barat.

**f. Persentase Tingkat Partisipasi Angkatan Kerja (TPAK)**

Semakin tingginya TPAK menunjukkan bahwa semakin tinggi pula pasokan tenaga kerja yang tersedia untuk produksi barang dan jasa. Hal tersebut yang mengindikasikan besarnya persentase penduduk usia kerja aktif secara ekonomi suatu daerah/wilayah. Berikut adalah persentase TPAK di Provinsi Maluku pada Gambar 6, sebagai berikut:



Gambar 6. Persentase Tingkat Partisipasi Angkatan Kerja (TPAK)

Data penelitian menunjukkan bahwa rata-rata persentase tingkat partisipasi angkatan kerja (TPAK) di Provinsi Maluku yaitu 65,79%, dimana Kabupaten Maluku Tenggara Barat memiliki persentase tertinggi sebesar 75,69% dan Kabupaten Seram Bagian Barat memiliki persentase terendah sebesar 58,25%. Berdasarkan Gambar 6, terlihat bahwa kabupaten/kota yang memiliki persentase TPAK yang lebih dari 58,25% dan kurang dari sama dengan 58,75% adalah Kabupaten Maluku Tengah, Kabupaten Seram Bagian Barat dan Kota Tual. Kabupaten/kota yang memiliki persentase TPAK yang lebih dari 58,75% dan kurang dari sama dengan 60,18% adalah kota Ambon. Kabupaten/kota yang memiliki persentase TPAK yang lebih dari 60,18% dan kurang dari sama dengan



70,26% adalah Kabupaten Buru Selatan, Kabupaten Buru dan Kabupaten Kepulauan Aru. Kabupaten/kota yang memiliki persentase TPAK yang lebih dari 70,26% dan kurang dari sama dengan 72,35% adalah Kabupaten Maluku Barat Daya dan Kabupaten Maluku Tenggara. Kabupaten/kota yang memiliki persentase TPAK yang lebih dari 72,35% dan kurang dari sama dengan 75,69% adalah Kabupaten Maluku Tenggara Barat.

### Pengujian Multikolinearitas

Uji multikolinearitas didasarkan atas asumsi bahwa model regresi tidak mengandung hubungan dependensi linier antara variabel prediktor. Jika Untuk mendeteksi adanya multikolinearitas dalam model dapat digunakan nilai *Variance Inflation Factor* (VIF). Hasil yang diperoleh dengan menggunakan software R adalah sebagai berikut :

Tabel 1. Nilai VIF

Variabel	VIF
X <sub>1</sub>	2,729
X <sub>2</sub>	1,707
X <sub>3</sub>	1,908
X <sub>4</sub>	1,083
X <sub>5</sub>	7,804

Berdasarkan Tabel 1, terlihat bahwa semua variabel independen memiliki nilai VIF kurang dari 10. Hal ini menunjukkan bahwa tidak terdapat kasus multikolinearitas, sehingga semua

Semua variabel independen dapat digunakan dalam memodelkan kemiskinan di Provinsi Maluku.

### Pemodelan Kemiskinan di Provinsi Maluku dengan Regresi Linier Berganda

Sebelum melakukan pemodelan kemiskinan dengan menggunakan pendekatan regresi spasial, maka akan dilakukan pemodelan kemiskinan dengan menggunakan regresi berganda. Dengan menggunakan software GeoDa, diperoleh

hasil estimasi parameter model regresi berganda sebagai berikut:

Tabel 2. Estimasi Parameter Regresi Berganda

Variabel	Koefisien	T	p-value
Constant	102,058	3,5065	0,0171
X <sub>1</sub>	0,23348	5,0208	0,0040
X <sub>2</sub>	-9,72971	-4,0738	0,0096
X <sub>3</sub>	1,77905	3,5208	0,0169
X <sub>4</sub>	-2,42616	-4,3302	0,0075
X <sub>5</sub>	-1,55064	-5,3612	0,0030

Berdasarkan Tabel 2, terlihat bahwa dengan menggunakan taraf nyata 0,05 maka kelima variabel prediktor yaitu persentase rumah tangga (RT) yang bahan bakar untuk memasaknya dari kayu (X<sub>1</sub>), persentase penduduk berumur 7 – 24 tahun yang tidak/belum pernah sekolah (X<sub>2</sub>), persentase penduduk berumur 7 – 24 tahun yang tidak sekolah lagi (X<sub>3</sub>), persentase tingkat pengangguran terbuka (TPT) (X<sub>4</sub>) dan persentase tingkat partisipasi angkatan kerja (X<sub>5</sub>) signifikan mempengaruhi persentase penduduk miskin di provinsi Maluku karena memiliki nilai p-value yang kurang dari 0,05. Sehingga model kemiskinan yang diperoleh dengan menggunakan regresi berganda adalah :

$$\hat{Y} = 102,058 + 0,23X_1 - 9,73X_2 + 1,78X_3 - 2,43X_4 - 1,55X_5$$

Berdasarkan model regresi berganda yang diperoleh dengan menggunakan metode OLS, maka koefisien determinasi yang dihasilkan dari model tersebut adalah 95,33% dan nilai AIC sebesar 52,46. Hal ini menunjukkan bahwa besarnya pengaruh kelima variabel prediktor tersebut terhadap persentase penduduk miskin di provinsi Maluku adalah sebesar 95,33%.

### Pemodelan Kemiskinan di Provinsi Maluku dengan Regresi Spasial

#### Identifikasi Awal Adanya Efek Spasial

Identifikasi ini untuk mengetahui apakah ada masalah dependensi spasial. Hal ini penting dilakukan untuk menentukan

tindakan selanjutnya, yaitu menentukan model spasial manakah yang akan digunakan untuk memodelkan kemiskinan. Dengan menggunakan software GeoDa diperoleh hasil sebagai berikut :

Tabel 3. Hasil Identifikasi Dependensi Spasial

No	Uji Spatial Dependensi	Statistik	p-value	Simpulan
1	Moran's I (error)	0,115	0,908	Terima $H_0$
2	Langrange Multiplier (lag)	3,972	0,046	Tolak $H_0$
3	Langrange Multiplier (error)	0,131	0,716	Terima $H_0$
4	Langrange Multiplier (SARMA)	4,089	0,129	Terima $H_0$

Moran's I bertujuan untuk mengetahui atau mengidentifikasi apakah ada spatial error atau tidak. Hipotesis yang digunakan yaitu sebagai berikut:

$H_0 : I = 0$  (tidak ada dependensi *spatial error*)

$H_1 : I \neq 0$  (ada dependensi *spatial error*)

Nilai *p-value* dari *Moran's I* sebesar 0,90823, dengan menggunakan taraf nyata 0,05 maka diperoleh kesimpulan terima  $H_0$  karena nilai *p-value* lebih dari 0,05, sehingga tidak ada dependensi spasial *error*. Selanjutnya dapat digunakan *Lagrange Multiplier* untuk endeteksi dependensi spasial secara lebih spesifik yaitu dependensi spasial dalam hal lag, *error*, atau keduanya (*lag* dan *error*).

#### a. Identifikasi dependensi lag

Uji *Lagrange Multiplier-Lag* digunakan untuk mengidentifikasi adanya keterkaitan antar wilayah. Hipotesis yang digunakan adalah:

$H_0 : \rho = 0$  (tidak ada dependensi spasial lag)

$H_1 : \rho \neq 0$  (ada dependensi spasial lag)

Berdasarkan Tabel 3, diperoleh nilai *p-value Lagrange Multiplier-Lag* sebesar 0,04624. Dengan menggunakan taraf nyata

0,05 maka diperoleh kesimpulan tolak  $H_0$  karena nilai *p-value* kurang dari 0,05, sehingga ada dependensi spasial lag. Oleh karena itu perlu dilanjutkan ke pembuatan *Spatial Lag Model*.

#### b. Identifikasi dependensi error

*Lagrange Multiplier-error* digunakan untuk mengidentifikasi adanya keterkaitan *error* antar wilayah. Hipotesis yang digunakan adalah :

$H_0 : \lambda = 0$  (tidak ada dependensi spasial *error*)

$H_1 : \lambda \neq 0$  (ada dependensi spasial *error*)

Berdasarkan Tabel 3, diperoleh nilai *p-value Lagrange Multiplier- Error* sebesar 0,71657. Dengan menggunakan taraf nyata 0,05 maka diperoleh kesimpulan terima  $H_0$  karena nilai *p-value* lebih dari 0,05, sehingga tidak ada dependensi spasial *error*. Oleh karena itu tidak perlu dilanjutkan ke pembuatan *Spatial Error Model*.

#### c. Identifikasi dependensi campuran (lag dan error)

*Lagrange Multiplier-SARMA* digunakan untuk mengidentifikasi adanya fenomena gabungan antara dependensi lag dengan dependensi *error* antar wilayah. Hipotesis yang digunakan adalah:

$H_0 : \rho, \lambda = 0$  (tidak ada dependensi spasial lag dan *error*)

$H_1 : \rho, \lambda \neq 0$  (ada dependensi spasial lag dan *error*)

Berdasarkan Tabel 3, diperoleh nilai *p-value Lagrange Multiplier-SARMA* sebesar 0,12939. Dengan menggunakan taraf nyata 0,05 maka diperoleh kesimpulan terima  $H_0$  karena nilai *p-value* lebih dari 0,05, sehingga tidak terdapat dependensi spasial campuran (lag dan *error*). Oleh karena itu tidak perlu dilanjutkan ke pembuatan *Spatial Autoregressive Moving Average* (SARMA).

#### d. Matriks Pembobot Spasial

Dalam penelitian ini akan digunakan dua matriks pembobot yaitu *Rook Contiguity* dan *Queen Contiguity*. Berdasarkan letak geografis tiap kabupaten/kota di provinsi Maluku (Gambar 3), diperoleh matriks pembobot yang sama antara *Rook Contiguity* dan *Queen Contiguity*. Hasil yang diperoleh adalah sebagai berikut:

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Pada Matriks  $C$  jumlah elemen setiap baris harus 1. Karena jumlah pada baris ketiga matriks  $C$  lebih dari 1, sehingga digunakan matriks pembobot spasial dalam bentuk standarisasi. Metode standarisasi yang digunakan yaitu standarisasi baris. Matriks pembobot spasial yang telah distandarisasi. Dengan menggunakan rumus standarisasi matriks  $W$  dengan elemen  $w_{ij}$  dinyatakan dengan  $w_{ij} = \frac{w_{ij}}{\sum_{j=1}^n w_{ij}}$

Sehingga diperoleh

$$w_{31} = \frac{1}{0+0+0+0+0+1+1+0+0+1+0} = 0,33$$

sehingga matriks standarisasi sebagai berikut:

$$W = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0,33 & 0,33 & 0 & 0 & 0,33 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

**e. Estimasi Parameter *Spatial Lag Model***

Setelah diperoleh matriks pembobot yang telah distandarisasi maka langkah selanjutnya melakukan estimasi parameter. Dengan menggunakan software GeoDa, diperoleh hasil estimasi parameter *spatial lag model* adalah sebagai berikut :

Tabel 4. Estimasi Parameter *Spatial Lag Model*

Variabel	Koefisien	Z	p-value
$\rho$	-0,23656	-2,7843	0,0053
Constant	66,1832	3,3290	0,0008
$X_1$	0,25711	10,146	0,0000
$X_2$	-6,76774	-3,9592	0,0000
$X_3$	1,80823	6,7848	0,0000
$X_4$	-1,28152	-2,5609	0,0104
$X_5$	-1,18058	-5,9692	0,0000

Berdasarkan Tabel 4, terlihat bahwa dengan menggunakan taraf nyata 0,05 maka kelima variabel prediktor yaitu persentase rumah tangga (rt) yang bahan bakar untuk memasaknya dari kayu ( $X_1$ ), persentase penduduk berumur 7 – 24 tahun yang tidak/belum pernah sekolah ( $X_2$ ), persentase penduduk berumur 7–24 tahun yang tidak sekolah lagi ( $X_3$ ), persentase tingkat pengangguran terbuka (TPT) ( $X_4$ ) dan persentase tingkat partisipasi angkatan kerja ( $X_5$ ) signifikan mempengaruhi persentase penduduk miskin di Provinsi Maluku karena memiliki nilai *p-value* yang kurang dari 0,05. Sehingga model kemiskinan yang diperoleh dengan menggunakan *spatial lag model* adalah :

$$\hat{y}_i = 66,18 - 0,24 \sum_{j=1}^n w_{ij}y_j + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_i$$

Dengan menggunakan matriks pembobot yang telah distandarisasi, maka diperoleh model untuk tiap kabupaten/kota di Provinsi Maluku yaitu:

1. Maluku Tenggara Barat

$$\hat{y}_{MTB} = 66,18 - 0,24y_{MALRA} + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_{MTB}$$

## 2. Maluku Tenggara

$$\hat{y}_{MALRA} = 66,18 - 0,24y_{TUAL} + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_{MALRA}$$

## 3. Maluku Tengah

$$\hat{y}_{MLTG} = 66,18 - 0,08y_{SBB} - 0,08y_{SBT} - 0,08y_{AMBON} + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_{MLTG}$$

## 4. Buru

$$\hat{y}_{BURU} = 66,18 - 0,24y_{BURSEL} + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_{BURU}$$

## 5. Kepulauan Aru

$$\hat{y}_{ARU} = 66,18 - 0,24y_{MALRA} + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_{ARU}$$

## 6. Seram Bagian Barat

$$\hat{y}_{SBB} = 66,18 - 0,24y_{MLTG} + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_{SBB}$$

## 7. Seram Bagian Timur

$$\hat{y}_{SBT} = 66,18 - 0,24y_{MLTG} + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_{SBT}$$

## 8. Maluku Barat Daya

$$\hat{y}_{MBD} = 66,18 - 0,24y_{MTB} + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_{MBD}$$

## 9. Buru Selatan

$$\hat{y}_{BRSL} = 66,18 - 0,24y_{BURU} + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_{BURSEL}$$

## 10. Ambon

$$\hat{y}_{AMQ} = 66,18 - 0,24y_{MLTG} + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_{AMQ}$$

## 11. Tual

$$\hat{y}_{TUAL} = 66,18 - 0,24y_{MALRA} + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_{TUAL}$$

Berdasarkan *spatial lag model* tiap kabupaten/kota di Provinsi Maluku, diperoleh bahwa :

1. Persentase penduduk miskin tiap kabupaten/kota akan bertambah sebesar 0,26 jika persentase rumah tangga yang bahan bakar untuk memasaknya berasal dari kayu bertambah sebesar satu satuan dan variabel yang lain dianggap konstan.
2. Persentase penduduk miskin tiap kabupaten/kota akan bertambah sebesar 6,77 jika persentase penduduk berumur 7 – 24 tahun yang tidak/belum pernah sekolah bertambah satu satuan dan variabel yang lain dianggap konstan.
3. Persentase penduduk miskin tiap kabupaten/kota akan bertambah sebesar 1,81 jika persentase penduduk berumur 7 – 24 tahun yang tidak sekolah lagi bertambah satu satuan dan variabel yang lain dianggap konstan.
4. Persentase penduduk miskin tiap kabupaten/kota akan berkurang sebesar 1,28 jika persentase tingkat pengangguran terbuka bertambah satu satuan dan variabel yang lain dianggap konstan.
5. Persentase penduduk miskin tiap kabupaten/kota akan berkurang sebesar 1,18 jika persentase tingkat partisipasi angkatan kerja bertambah satu satuan dan variabel yang lain dianggap konstan.

Berdasarkan *spatial lag model* yang diperoleh, maka *koefisien determinasi* yang dihasilkan dari model tersebut adalah 97,18% dan nilai AIC sebesar 49,17. Hal ini menunjukkan bahwa besarnya pengaruh kelima variabel prediktor tersebut terhadap persentase penduduk miskin di Provinsi Maluku adalah sebesar 97,18%.

## Perbandingan Model Regresi Berganda dan *Spatial Lag Model*

Setelah diperoleh model regresi berganda dan *spatial lag model*, maka selanjutnya akan dipilih model terbaik

untuk menggambarkan kemiskinan di Provinsi Maluku. Kriteria yang digunakan untuk memilih model terbaik yaitu  $R^2$  dan AIC. Hasilnya disajikan dalam tabel berikut ini :

Tabel 5. Nilai  $R^2$  dan AIC

Model	$R^2$	AIC
Regresi Berganda	95,33%	52,46
Spatial Lag Model dengan Matriks Pembobot : WRook/Queen Contiguity	97,18%	49,17

Berdasarkan Tabel 5, terlihat bahwa nilai  $R^2$  dan AIC dari model regresi berganda adalah 95,33% dan 52,46; sedangkan nilai  $R^2$  dan AIC dari *spatial lag model* adalah 97,18% dan 49,17. Hal ini menunjukkan bahwa model terbaik untuk menggambarkan kemiskinan di provinsi Maluku yaitu *spatial lag model* karena memiliki nilai  $R^2$  tertinggi maupun nilai AIC terendah.

#### Pengujian Asumsi Residual dari *Spatial Lag Model*

Model regresi akan dapat dijadikan alat estimasi yang baik dan tidak bias jika telah dilakukan beberapa pengujian asumsi residual. Hasil yang diperoleh adalah sebagai berikut :

#### Uji Heteroskedastisitas

Asumsi yang harus dipenuhi oleh suatu model adalah memiliki variansi *error* yang konstan/tetap, yang lebih dikenal dengan istilah homoskedastisitas. Untuk mendeksi adanya heteroskedastisitas akan digunakan uji *Breusch-Pagan*. Hasil diperoleh sebagai berikut :

Tabel 6. Uji *Breusch-Pagan* Untuk Deteksi Heteroskedastisitas

Statistic	DF	p-value
1,7554	5	0,88185

Berdasarkan Tabel 6, terlihat bahwa nilai *p-value* dari uji *Breusch-Pagan* sebesar 0,88185. Dengan menggunakan taraf nyata 5% maka diperoleh kesimpulan terima  $H_0$ . Sehingga variansi residual bersifat homoskedastisitas.

#### Uji Autokorelasi

Salah satu asumsi yang harus dipenuhi oleh model regresi yaitu residualnya harus bersifat independen untuk setiap observasi. Untuk memeriksa suatu model memenuhi asumsi ini dapat menggunakan uji Run Test. Hasil diperoleh sebagai berikut :

Tabel 7. Uji Run Test Untuk Autokorelasi Residual

	Unstandardized Residual
Test Value <sup>a</sup>	0,17
Total Cases	11
Number of Runs	7
Z	0,029
Asymp. Sig. (2-tailed)	0,977

Berdasarkan Tabel 7, terlihat bahwa nilai *p-value* lebih dari taraf nyata 0,05 sehingga terima  $H_0$ . Hal ini menunjukkan bahwa tidak terdapat masalah autokorelasi.

#### Uji Normalitas

Uji normalitas merupakan salah satu uji yang bertujuan untuk mendeteksi residual berdistribusi normal atau tidak. Salah satu metode yang digunakan untuk mendeteksi kenormalan residual yaitu metode *Kolmogorov-Smirnov*. Hasil diperoleh sebagai berikut :

Tabel 8. Uji Kolmogorov-Smirnov Untuk Residual

Statistic	Df	Sig.
0,142	11	0,200

Berdasarkan Tabel 8, terlihat bahwa nilai *Sig.* dari uji *Kolmogorov-Smirnov* lebih dari taraf signifikansi 0,05. Hal ini menunjukkan bahwa residual berdistribusi normal.

Berdasarkan hasil pengujian asumsi, terlihat bahwa model yang dihasilkan telah memenuhi asumsi-asumsi yang ada, sehingga dapat disimpulkan bahwa model yang dihasilkan dapat digunakan untuk menggambarkan ersentase penduduk miskin di Provinsi Maluku.

## KESIMPULAN

Berdasarkan hasil dan pembahasan, maka kesimpulan yang dapat diambil dari penelitian ini yaitu:

- Rata-rata persentase penduduk miskin di provinsi Maluku yaitu 22,71%, dimana Kabupaten Maluku Barat Daya memiliki persentase tertinggi sebesar 31,01% dan Kota Ambon memiliki persentase terendah sebesar 4,64%.
- Faktor-faktor yang signifikan mempengaruhi kemiskinan di Provinsi Maluku dengan menggunakan regresi spasial adalah persentase rumah tangga (RT) yang bahan bakar untuk memasaknya dari kayu ( $X_1$ ), persentase penduduk berumur 7 – 24 tahun yang tidak/belum pernah sekolah ( $X_2$ ), persentase penduduk berumur 7 – 24 tahun yang tidak sekolah lagi ( $X_3$ ), persentase tingkat pengangguran terbuka (TPT) ( $X_4$ ) dan persentase tingkat partisipasi angkatan kerja ( $X_5$ )
- Model terbaik yang diperoleh untuk memodelkan kemiskinan di Provinsi Maluku yaitu *Spatial Lag Model* :

$$\widehat{y}_i = 66,18 - 0,24 \sum_{j=1}^n w_{ij}y_j + 0,26X_1 + 6,77X_2 + 1,81X_3 - 1,28X_4 - 1,18X_5 + \varepsilon_i$$

## DAFTAR PUSTAKA

- Akaike, H. (1974). A New Look at Statistical Model Identification. *IEEE Transaction on Automatic Control*, Vol. 19, 328-347
- Anselin, L. (1988). *Spatial Econometrics : Methods and Models*. Dordrecht: Kluwer Academic Publishers.
- Anton, H. (1992). *Dasar-dasar Aljabar Linier* oleh Syarifudin, Mike Damayanti, Yayan Wulandari ; Tangerang; Binapura Aksara
- Aulele, S. N. (2014). Pendekatan Model Geographically Weighted Regression untuk Menentukan Faktor-faktor yang Mempengaruhi Jumlah Rumah Tangga Miskin di Pulau Buru. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 8(2), 53-57.

- Badan Pusat Statistik dan Depsos/Departemen Sosial (2002), *Penduduk Fakir Miskin Indonesia 2002*, Jakarta: BPS
- Bain, L. J., & Engelhard, M. (1992). *Introduction to Probability and Mathematical Statistics*. California: Duxbury Press.
- BPS Provinsi Maluku. 2017. *Profil Kemiskinan di Maluku tahun 2017*. Badan Pusat Statistik Provinsi Maluku: Ambon.
- Draper, N. R., & Smith, H. (1992). *Analisis Regresi Terapan*. Jakarta: PT. Gramedia Pustaka Utama.
- Johnson, R.A. dan Wichern, D.W., 2007, *Applied Multivariat Statistical Analysis*, New Jersey, Prentice Hall.
- Laswinia, V. D., & Chamid, M. S. (2016). Analisis Pola Hubungan Persentase Penduduk Miskin dengan Faktor Lingkungan, Ekonomi, dan soial di Indonesia, Menggunakan Regresi Spasial. *Jurnal Sains dan Seni ITS* Vol 5 No 2.
- Leleury, Z. A., & Tomasouw, B. P. (2019). Pengelompokan dan Pemetaan Karakteristik Kemiskinan di Kabupaten Maluku Barat Daya Provinsi Maluku dengan menggunakan Self-Organizing Map dan Analisis Biplot. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 13(2), 093-106

# PEMBENTUKAN PORTOFOLIO SAHAM MENGGUNAKAN KLASTERING TIME SERIES K-MEDOID DENGAN UKURAN JARAK DYNAMIC TIME WARPING

La Gubu<sup>1</sup>, Dedi Rosadi<sup>2</sup>, and Abdurakhman<sup>3</sup>

<sup>1,2,3</sup>Departemen Matematika Universitas Gadjah Mada, Yogyakarta-Indonesia

<sup>1</sup>Jurusan Matematika Universitas Halu Oleo, Kendari-Indonesia

e-mail: <sup>1</sup>lagubu2014@gmail.com, <sup>2</sup>dedirosadi@gadjahmada.edu, <sup>3</sup>rachmanstat@ugm.ac.id

## Abstrak

Pada penelitian ini akan disajikan pembentukan portofolio saham dengan *preprocessing* data menggunakan klastering *time series* dengan ukuran jarak *Dynamic Time Warping* (DTW). Pertama-tama saham-saham dikelompokkan ke dalam beberapa klaster menggunakan klastering *time series Partitioning Around Medoids* (PAM) berdasarkan ukuran jarak DTW. Setelah proses klastering, saham dipilih untuk mewakili masing-masing klaster untuk membangun portofolio optimum. Saham yang dipilih dari masing-masing klaster merupakan saham yang memiliki *Sharpe ratio* tertinggi. Portofolio optimal ditentukan dengan menggunakan tiga model portofolio, yaitu: model portofolio *Mean-Variance* (MV) klasik, model portofolio MV *robust Fast Minimum Covariance Determinant* (FMCD) dan model portofolio *robust Scale* (S). Dengan menggunakan prosedur ini, dapat diperoleh portofolio optimum secara efisien bila ada banyak saham yang terlibat dalam proses pembentukan portofolio. Untuk mengukur kinerja portofolio yang terbentuk digunakan *Sharpe ratio*. Hasil kajian empiris menunjukkan bahwa kinerja portofolio yang dihasilkan dengan menggunakan klastering *time series* PAM dengan ukuran disimilaritas jarak DTW yang dikombinasikan dengan model portofolio MV klasik mengungguli kinerja portofolio yang dihasilkan kombinasi dengan model yang lain.

**Kata kunci:** klastering, *k-medoids*, portofolio saham, estimasi *robust*, kinerja portofolio

## Abstract

*This research will present the formation of stock portfolios by preprocessing data using time-series clustering with a distance measure of Dynamic Time Warping (DTW). First, stocks are grouped into several clusters using the Partitioning Around Medoids (PAM) time series cluster based on the DTW distance measure. After the clustering process, stocks are selected to represent each cluster to build the optimum portfolio. The stock selected from each cluster is the one with the highest Sharpe ratio. The optimal portfolio is determined using three portfolio models, namely: the classic Mean-Variance (MV) portfolio model, the FMCD robust MV portfolio model, and the S robust portfolio model. Using this procedure, an optimum portfolio can be obtained efficiently if there are many stocks involved in the portfolio formation process. Sharpe ratio is used to measure the performance of the portfolios. The results of the empirical study show that the portfolio performance generated using the PAM time series clustering with DTW distance dissimilarity measure combined with the classic MV portfolio model outperforms the resulting portfolio performance in combination with other models.*

**Keywords:** clustering, *k-medoid*, stock portfolio, robust estimation, portfolio performance

## PENDAHULUAN

Hakikat pembentukan portofolio adalah mengalokasikan modal pada berbagai alternatif investasi, sehingga risiko investasi dapat diminimumkan (Elton, dkk., 2014). Dalam memilih portofolio, *investor* berusaha memaksimalkan keuntungan dari investasi dengan tingkat risiko tertentu yang diterima. Atau dalam pengertian lain, *investor* berusaha meminimalkan risiko yang dihadapi pada tingkat keuntungan tertentu. Portofolio yang dapat mencapai tujuan tersebut disebut portofolio efisien (Reilly dan Brown, 2012). Apabila seorang *investor* memiliki beberapa pilihan portofolio yang efisien, maka portofolio optimal yang akan dipilihnya. Portofolio yang optimal adalah portofolio yang dapat memaksimalkan preferensi *investor* sehubungan dengan tingkat keuntungan dan risiko.

Pembentukan portofolio merupakan masalah yang menarik dan menantang dalam studi pemodelan ekonomi dan keuangan. Masalah utama dalam pembentukan portofolio adalah memaksimalkan return sekaligus meminimalkan risiko, sehingga dapat dirumuskan menjadi masalah optimasi (Supandi, 2017). Upaya pertama dalam memodelkan masalah optimasi ini telah dilakukan oleh Markowitz pada tahun 1952. Markowitz (1952) menggunakan ukuran-ukuran statistik dari data historis harga saham untuk membangun portofolio optimal. Dalam model ini, return saham direpresentasikan sebagai mean dari data sedangkan risiko direpresentasikan sebagai variansi, sehingga model ini disebut model portofolio *Mean-Variance* (MV).

Studi untuk mengembangkan dan memperluas model MV lebih lanjut telah dilakukan oleh banyak peneliti. Pengembangan dan perluasan model tersebut dilakukan untuk menyesuaikan model agar lebih adaptif dan *robust* terhadap kondisi pasar keuangan yang terus berubah, serta membangun cara yang lebih efektif dan efisien untuk membangun portofolio yang optimal. Salah satu cara yang dilakukan adalah dengan metode

klustering. Penelitian tentang pembentukan portofolio dengan menggunakan klustering telah dilakukan oleh beberapa peneliti, antara lain oleh Guan dan Jiang (2007), Tola dkk (2008), Chen dan Huang (2009), Nanda dkk (2010), Long dkk (2014), dan Gubu, dkk (2019). Dalam pendekatan ini, saham dipisahkan menjadi beberapa kelompok (klaster), kemudian dipilih perwakilan dari masing-masing kelompok untuk membentuk portofolio optimal dengan menggunakan model portofolio MV.

Masalah utama model portofolio MV adalah bahwa vektor mean dan matriks variansi-kovariansi harus diestimasi dari data yang dapat sangat volatil. Estimasi parameter dapat dilakukan dengan menggunakan berbagai pilihan teknik estimasi, yang tentunya akan mengandung kesalahan estimasi. Sebagai input yang sangat penting dalam pembentukan model portofolio MV, kesalahan estimasi akan berpengaruh signifikan terhadap hasil pembentukan portofolio optimal. Beberapa penelitian terkait dengan kesalahan estimasi dan hubungannya dengan pembentukan portofolio optimal telah dilakukan oleh Best dan Grauer (1991), Chopra dan Ziemba (1993), dan Ceria dan Stubbs (2006). Studi-studi tersebut menyimpulkan bahwa meskipun model MV didukung oleh teori yang kuat dan relatif mudah dalam komputasi, model MV menunjukkan beberapa kelemahan, diantaranya portofolio optimal yang dihasilkan oleh model ini belum terdiversifikasi dengan baik. Selain itu, model MV juga sangat sensitif terhadap perubahan parameter input yaitu vektor mean dan matriks variansi-kovariansi.

Oleh karena itu beberapa peneliti telah membangun suatu portofolio *robust* (Demiguel dan Nogales, 2009), yaitu portofolio yang dapat mengurangi kesalahan estimasi vektor mean dan matriks variansi-kovariansi pada model portofolio MV. Salah satu pendekatan standar dalam membentuk portofolio *robust* optimal adalah melalui pendekatan estimasi *robust*. Beberapa studi tentang pembentukan portofolio optimal menggunakan estimasi



*robust* telah dilakukan oleh Lauprete (2001), Vaz-de Melo dan Camara (2005), Welsh dan Zhou (2007), DeMiguel dan Nogales (2008), dan Supandi (2017). Perbedaan utama antara penelitian-penelitian ini terletak pada estimasi *robust* yang digunakan dalam pembentukan portofolio optimal.

Berdasarkan penelitian-penelitian sebelumnya, disimpulkan bahwa, untuk meningkatkan efisiensi proses pembentukan portofolio, perlu mengelompokkan saham-saham ke dalam klaster-klaster, kemudian memilih saham representasi dari masing-masing klaster dan menentukan bobot masing-masing saham representasi.

Pada penelitian ini, saham-saham dikelompokkan ke dalam klaster-klaster menggunakan klastering *time series Partitioning Around Medoids* (PAM) dengan ukuran jarak *Dynamic Time Warping* (DTW).

Pemilihan PAM sebagai analisis klaster yang digunakan dengan pertimbangan bahwa analisis klaster PAM lebih *robust* dibandingkan dengan analisis klaster yang lain seperti *k-mean*. Pada analisis klaster PAM, pusat klaster adalah *medoid* dari data dalam suatu klaster, sehingga jika ada *outlier* dalam data maka analisis klaster PAM akan lebih *robust* dibandingkan dengan analisis klaster *k-mean*.

Ukuran jarak DTW juga disebut sebagai *non-linear sequence alignment*, sehingga ukuran jarak ini lebih realistis untuk digunakan dalam mengukur disimilaritas suatu pola dibandingkan menggunakan ukuran jarak yang lain seperti jarak Euclidean, Manhattan, Minkowski dan lain-lain (Berndt dan Clifford, 1994). Data yang diolah selalu berada dalam kawasan waktu, sehingga rentetan data yang kita miliki dianggap bervariasi terhadap waktu. Atas dasar ini maka ukuran jarak DTW dipilih untuk digunakan dalam penelitian ini.

Bobot portofolio ditentukan dengan menggunakan tiga model portofolio, yaitu model portofolio MV klasik, model portofolio MV *robust* FMCD dan model

portofolio MV *robust* S. Untuk mengukur kinerja portofolio yang terbentuk dengan menggunakan tiga model portofolio digunakan *Sharpe ratio*.

## METODOLOGI

### Model Portofolio Mean-Variance

Model portofolio *Mean-Variance* (MV) untuk membentuk portofolio optimal didasarkan pada *trade-off* antara return dan risiko portofolio yang direpresentasikan sebagai mean dan variansi saham (Markowitz, 1952). Oleh karena itu, model ini dapat dirumuskan sebagai masalah optimasi berikut (Supandi, 2017):

$$\max_{\mathbf{w}} \mathbf{w}'\boldsymbol{\mu} - \frac{\gamma}{2} \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \quad (1)$$

$$\mathbf{w}'\mathbf{e} = 1 \quad (2)$$

dengan  $\mathbf{w}$  menyatakan bobot portofolio,  $\boldsymbol{\mu}$  adalah vektor mean,  $\boldsymbol{\Sigma}$  adalah matriks variansi-kovariansi,  $\mathbf{e}$  adalah matriks kolom dengan semua elemennya adalah 1 dan  $\gamma \geq 0$  adalah parameter penghindaran risiko (*risk aversion*), yaitu ukuran relatif penghindaran risiko. Model portofolio MV dipilih oleh *investor* dengan menggunakan kriteria fungsi utilitas. Dengan asumsi bahwa sikap *investor* adalah *risk averse*, *investor* akan memilih portofolio yang memberikan utilitas maksimum yang diharapkan  $E(U)$ . Jadi, portofolio *mean-variance* setara dengan

$$\max E(U(\mathbf{w})) \quad (3)$$

$$\mathbf{w}'\mathbf{e} = 1 \quad (4)$$

Diberikan modal awal  $W_0$ , di bawah portofolio dengan vektor bobot  $\mathbf{w}$ , pada akhir periode, modal menjadi  $W_0(1 + R_p)$ , di mana  $R_p$  adalah variabel random return portofolio (Supandi, 2017). Fungsi utilitas  $W_0(1 + R_p)$  adalah  $U(W_0(1 + R_p))$ . Selanjutnya, Supandi (2017) menyatakan bahwa fungsi utilitas dapat diperluas dengan menggunakan pendekatan deret Taylor orde dua, yaitu

$$U(W_0(1 + R_p)) = U(W_0) + W_0 U'(W_0) R_p + \frac{1}{2} W_0^2 U''(W_0) R_p^2 + O(R_p^3) \quad (5)$$

Mengambil nilai harapan persamaan (5) diperoleh

$$E\left(U(W_0(1 + R_p))\right) = U(W_0) + W_0 U'(W_0) \left(\mu_p + \frac{1}{2} W_0 \frac{U''(W_0)}{U'(W_0)} \sigma_p^2\right) \quad (6)$$

Berdasarkan pendekatan tersebut, memaksimalkan fungsi utilitas yang diharapkan sama dengan memaksimalkan

$$A = \mu_p - \frac{1}{2} \gamma \sigma_p^2 \quad (7)$$

dengan  $\gamma = -W_0 \frac{U''(W_0)}{U'(W_0)}$  menyatakan ukuran relatif penghindaran risiko. Telah diketahui bahwa  $\mu_p = \mathbf{w}'\boldsymbol{\mu}$  dan  $\sigma_p^2 = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$ , sehingga persamaan (7) dapat ditulis dalam bentuk

$$B = \mathbf{w}'\boldsymbol{\mu} - \frac{\gamma}{2} \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \quad (8)$$

Sehingga persamaan (1) ekuivalen dengan persamaan (8) ditambah kendala  $\mathbf{w}'\mathbf{e} = 1$ .

Permasalahan optimasi pada persamaan (1) dan (2) dapat diselesaikan dengan menggunakan metode Lagrange. Pertama, bentuk fungsi Lagrange:

$$L = \mathbf{w}'\boldsymbol{\mu} - \frac{1}{2} \gamma \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} + \lambda(\mathbf{w}'\mathbf{e} - 1) \quad (9)$$

Berdasarkan teorema Kuhn-Tucker (Winston dan Goldberg, 2004), syarat perlu persamaan (9) mencapai nilai optimum adalah:

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \quad (10)$$

$$\frac{\partial L}{\partial \lambda} = 0 \quad (11)$$

Dari persamaan (9), (10) dan (11) diperoleh

$$\mathbf{w} = \frac{\boldsymbol{\Sigma}^{-1}}{\gamma} (\boldsymbol{\mu} + \lambda \mathbf{e}) \quad (12)$$

dan

$$\mathbf{e}'\mathbf{w} = 1 \quad (13)$$

Substitusi persamaan (12) ke persamaan (13) diperoleh

$$\lambda = \gamma(\mathbf{e}'\boldsymbol{\Sigma}^{-1}\mathbf{e})^{-1} - (\mathbf{e}'\boldsymbol{\Sigma}^{-1}\mathbf{e})^{-1} \mathbf{e}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \quad (14)$$

Substitusi persamaan (14) pada persamaan (12) menghasilkan:

$$\mathbf{w} = \frac{1}{\gamma} (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{e}(\mathbf{e}'\boldsymbol{\Sigma}^{-1}\mathbf{e})^{-1}\mathbf{e}'\boldsymbol{\Sigma}^{-1})\boldsymbol{\mu} + \boldsymbol{\Sigma}^{-1}\mathbf{e}(\mathbf{e}'\boldsymbol{\Sigma}^{-1}\mathbf{e})^{-1} \quad (15)$$

Persamaan (15) menunjukkan bahwa bobot portofolio ( $\mathbf{w}$ ) bergantung pada input vektor mean  $\boldsymbol{\mu}$  dan matiks variansi-kovariansi  $\boldsymbol{\Sigma}$ .

### Klastering Time Series

Masalah klastering *time series* terjadi ketika kita mengamati himpunan *time series* dan kita ingin mengaturnya ke dalam kelompok atau kluster terpisah (Alonso,

2006). Pada bagian ini akan disajikan secara singkat tentang hal-hal yang terkait dengan analisis kluster *time series* yang akan digunakan dalam penelitian ini.

### Ukuran Disimilaritas

Ukuran disimilaritas (*dissimilarity measure*) merupakan hal terpenting dalam melakukan analisis kluster. Banyak ukuran disimilaritas antara *time series* telah diajukan dalam literatur. Dalam penelitian ini ukuran disimilaritas yang digunakan adalah jarak *Dynamic Time Warping* (DTW).

Jarak DTW telah dipelajari secara mendalam oleh Sankoff dan Kruskal (1983) dan digunakan untuk menemukan pola dalam deret waktu oleh Berndt dan Clifford (1994). Jarak DTW bertujuan untuk menemukan pemetaan  $r$  antara deret waktu tersebut sehingga ukuran jarak antara pengamatan berpasangan  $(X_{a_i}, Y_{b_i})$  adalah minimum. Montero dan Vilar (2014) memberikan definisi jarak DTW sebagai berikut.

$$d_{DTW}(\mathbf{X}_T, \mathbf{Y}_T) = \min_{r \in M} (\sum_{i=1, \dots, m} |X_{a_i} - Y_{b_i}|) \quad (16)$$

Lebih lanjut, Montero dan Vilar (2014) menyatakan bahwa jarak DTW memungkinkan untuk mengenali bentuk yang serupa, bahkan dengan adanya transformasi sinyal seperti pergeseran dan/atau penskalaan.

### k-Medoids Clustering

Klastering *k-medoids* adalah teknik klastering yang mirip dengan teknik klastering *k-means* untuk mempartisi kumpulan data (*data set*) menjadi  $k$  kelompok atau kluster. Dalam pengelompokan *k-medoids*, setiap kluster diwakili oleh salah satu titik data dalam kluster. Titik-titik ini disebut *medoids* dari kluster. Istilah *medoid* mengacu pada entitas dalam sebuah kluster yang memiliki perbedaan rata-rata minimal antara titik tersebut dan semua anggota kluster lainnya. Sebagai catatan, dalam klastering *k-means*, pusat kluster ditentukan sebagai nilai rata-rata dari semua titik data kluster. *K-medoids* adalah alternatif yang *robust* untuk klastering *k-means*. Ini berarti bahwa klastering *k-medoids* kurang sensitif

terhadap pencilaan dibandingkan dengan klastering *k-means*, karena metode klastering tersebut menggunakan *medoid* sebagai pusat cluster. *Partitioning Around Medoids* (PAM) adalah pendekatan klastering yang paling umum untuk *k-medoids* (Kaufman dan Rousseeuw, 1990).

Algoritma PAM didasarkan pada pencarian *k* objek perwakilan (*representative*) atau *medoid* di antara pengamatan kumpulan data (*data set*). Setelah menemukan satu set *k medoid*, klaster dibangun dengan menetapkan setiap observasi ke *medoid* terdekat. Selanjutnya, setiap *medoid* yang dipilih dan setiap titik data *non-medoid* ditukar (*swapped*) dan fungsi tujuan dihitung. Fungsi objektif adalah jumlah disimilaritas semua objek ke *medoid* terdekatnya.

Langkah pertukaran (*swap*) mencoba meningkatkan kualitas pengelompokan dengan menukar objek yang dipilih sebagai *medoid* dan objek *non-medoid*. Jika nilai fungsi objektif dapat dikurangi dengan menukar objek yang dipilih sebagai *medoid* dengan objek *non-medoid*, maka pertukaran dilakukan. Hal ini terus dilakukan hingga fungsi objektif tidak dapat diturunkan lagi. Tujuannya adalah untuk menemukan objek perwakilan yang meminimalkan jumlah disimilaritas pengamatan ke objek perwakilan terdekat mereka.

Secara terperinci, Kaufman dan Rousseeuw (1990) menguraikan langkah-langkah algoritma analisis klaster PAM sebagai berikut:

1. Pilih *k* objek untuk menjadi *medoid*;
2. Hitung matriks disimilaritas menggunakan persamaan (16);
3. Tetapkan setiap objek ke *medoid* terdekatnya;
4. Untuk setiap pencarian klaster jika salah satu objek klaster menurunkan koefisien ketidaksamaan rata-rata, pilih entitas yang paling banyak menurunkan koefisien ini sebagai *medoid* untuk klaster ini;
5. Jika setidaknya satu *medoid* telah berubah, lanjutkan ke (3), jika tidak, akhiri algoritma.

### **Sharpe Ratio**

Setelah klaster-klaster terbentuk, selanjutnya dilakukan penilaian terhadap kinerja masing-masing saham dalam setiap klaster dengan menggunakan Sharpe *ratio* (SR). Sharpe *ratio* atau indeks Sharpe adalah ukuran dari *excess return* (atau *risk premium*) per unit risiko dalam aset (Sharpe, 1994). Sharpe *ratio* digunakan untuk mengkarakterisasi seberapa baik return aset mengkompensasi *investor* untuk risiko yang diambil. Lebih lanjut Sharpe (1994) menyatakan bahwa SR dihitung dengan membandingkan selisih antara return saham (*R*) dan *return risk-free rate* (*R<sub>f</sub>*) dengan standar deviasi return saham (*σ*) atau dapat ditulis sebagai berikut:

$$SR = \frac{R - R_f}{\sigma} \quad (17)$$

Secara umum dapat dikatakan bahwa semakin besar nilai Sharpe *ratio* suatu saham maka kinerja saham semakin baik. Sharpe *ratio* juga dapat digunakan untuk mengukur kinerja portofolio. Jika Sharpe *ratio* digunakan untuk mengukur kinerja portofolio maka return dan risiko yang digunakan adalah return dan risiko portofolio.

### **Pemilihan Portofolio Optimum dengan Menggunakan Estimasi Robust**

Ada dua terminologi penting dalam pembentukan portofolio, yaitu portofolio efisien dan portofolio optimum. Portofolio efisien adalah portofolio dengan return tertinggi pada risiko tertentu, atau portofolio dengan risiko terendah pada return tertentu (Elton, dkk., 2014). Investor perlu mempertimbangkan dan menentukan sekuritas apa saja yang membentuk portofolio dan dapat mencapai efisiensi maksimal. Indikator portofolio efisien adalah:

1. Mampu memberikan *expected* return terbesar dengan risiko yang sama. \
2. Mampu memberi risiko terkecil dengan *expected* return yang sama.

Penentuan portofolio efisien dilakukan dengan cara memilih tingkat *expected* return tertentu dan meminimumkan risikonya, atau menentukan tingkat risiko tertentu dan kemudian memaksimumkan *expected* returnnya. Portofolio efisien merupakan

portofolio yang baik, tetapi bukan yang terbaik. Portofolio yang terbaik adalah portofolio yang optimal. Portofolio efisien hanya mempunyai satu dari faktor terbaik, yaitu faktor *expected* return atau faktor risikonya. Sementara, portofolio yang optimal adalah portofolio yang memiliki kombinasi *expected* return dan risiko yang terbaik (Elton, dkk., 2014).

Portofolio MV klasik menjadi tidak efektif ketika dihadapkan pada kondisi data return yang tidak memenuhi asumsi distribusi normal multivariat, karena estimasi klasik dari mean dan variansi tidak *robust* dan sangat dipengaruhi oleh pengamatan-pengamatan yang menyimpang (*outlier*), (Maronna dkk, 2006). Statistika *robust* berkaitan dengan membangun prosedur statistik yang stabil ketika ada bagian dari data yang tidak sesuai dengan distribusi yang diasumsikan atau terdapat penyimpangan dari model.

Pada penelitian ini, vektor mean dan matriks variansi-kovariansi diestimasi dengan menggunakan metode estimasi *robust Fast Minimum Covariance Determinant* (FMCD) dan metode estimasi *robust Scale* (S). Berikut akan disajikan secara sekilas prosedur metode estimasi *robust* FMCD dan metode estimasi *robust* S.

### Metode Estimasi *Robust* FMCD

Estimasi *Minimum Covariance Determinant* (MCD) bertujuan untuk mencari estimasi *robust* berdasarkan  $h$  pengamatan dari total pengamatan ( $n$ ), dimana matriks kovariansi mempunyai determinan paling kecil. Estimasi MCD merupakan pasangan  $\hat{\mu} \in \mathbb{R}^p$  dan  $\hat{\Sigma}$  adalah matriks definit positif simetris berdimensi  $p \times p$  dari suatu sub ruang sampel berukuran  $h$  pengamatan, di mana  $\frac{(n+p+1)}{2} \leq h \leq n$  dengan

$$\hat{\mu} = \frac{1}{h} \sum_{i=1}^h \mathbf{r}_i \quad (18)$$

dengan  $\mathbf{r}_i$  return saham ke- $i$ ,  $i = 1, \dots, h$ .

Estimasi matriks kovariansi dapat diperoleh dengan menyelesaikan persamaan berikut ini:

$$\hat{\Sigma} = \frac{1}{h} \sum_{i=1}^h (\mathbf{r}_i - \hat{\mu})(\mathbf{r}_i - \hat{\mu})' \quad (19)$$

Perhitungan MCD bisa menjadi sangat rumit apabila dimensi data semakin besar, hal ini dikarenakan pada metode ini harus memeriksa semua kemungkinan himpunan bagian  $h$  dari sejumlah  $n$  data. Oleh karena itu Rousseeuw dan Van Driessen (1999) menemukan algoritma perhitungan yang lebih cepat untuk menghitung MCD yang disebut *Fast MCD* (FMCD). Metode FMCD berdasarkan teorema C-Step yang dijelaskan berikut ini.

**Teorema 1** (Rousseeuw dan Driessen, 1999) *Jika  $H_1$  adalah himpunan bagian berukuran  $h$  yang diambil dari data berukuran  $n$ , maka statistik sampel adalah:*

$$\hat{\mu}^1 = \frac{1}{h} \sum_{i \in H_1} \mathbf{r}_i \quad (20)$$

$$\hat{\Sigma}^1 = \frac{1}{h} \sum_{i \in H_1} (\mathbf{r}_i - \hat{\mu}^1)(\mathbf{r}_i - \hat{\mu}^1)' \quad (21)$$

*Jika  $|\hat{\Sigma}^1| > 0$  maka jarak  $d_i = (\mathbf{r}_i; \hat{\mu}^1, \hat{\Sigma}^1)$ . Selanjutnya tentukan  $H_2$  himpunan bagian yang terdiri dari  $h$  pengamatan dengan jarak terkecil  $d_i$ , yaitu  $\{d_1(i) | i \in H_2\} = \{(d_1)_1, \dots, (d_1)_h\}$  dengan  $(d_1)_1 \leq (d_1)_2 \leq \dots \leq (d_1)_n$  merupakan jarak yang berurutan. Berdasarkan  $H_2$  dengan menggunakan persamaan (20) dan (21), diperoleh*

$$|\hat{\Sigma}^2| \leq |\hat{\Sigma}^1| \quad (22)$$

*Ekspresi (22) akan sama apabila  $\hat{\mu}^1 = \hat{\mu}^2$  dan  $\hat{\Sigma}^1 = \hat{\Sigma}^2$ .*

Teorema C-Step dilakukan berulang kali sampai  $|\hat{\Sigma}_{baru}| > 0$  atau  $|\hat{\Sigma}_{baru}| = |\hat{\Sigma}_{lama}|$ .

### Metode Estimasi *Robust* S

Estimasi ini diperkenalkan pertama kali oleh Rousseeuw and Yohai (1984) yang kemudian dikembangkan lagi oleh Lopuhaa (1989) dan Davies (1987).

**Definisi 1** (Davies, 1987) *Diberikan  $\{\mathbf{r}_i, i = 1, \dots, n\}$  adalah himpunan data di  $\mathbb{R}^p$  dan diberikan  $\mathcal{P}_p$  adalah himpunan semua matriks simetrik definit positif berukuran  $p \times p$ . Estimasi S untuk ukuran lokasi  $\hat{\mu} \in \mathbb{R}^p$  dan dispersi  $\hat{\Sigma}(R) \in \mathcal{P}_p$  adalah setiap pasangan yang meminimumkan  $|\Sigma|$  dengan kendala  $\frac{1}{n} \sum_{i=1}^n \rho[(\mathbf{r}_i - \mu)' \Sigma^{-1} (\mathbf{r}_i - \mu)]^{1/2} = b_0$*

*dengan  $\rho$  adalah fungsi kerugian (loss function), dan  $b_0$  adalah suatu*

*konstanta*. Konstanta ini harus ditentukan dengan tepat karena nilai ini berpengaruh terhadap hasil estimasi. Apabila distribusi data tidak diketahui maka dipilih  $b_0 = E\{\rho\|\mathbf{r}\|\}$ .

Estimator S dapat diperoleh dengan menyelesaikan persamaan berikut ini:

$$\frac{1}{n} \sum_{i=1}^n u(d_i)(\mathbf{r}_i - \boldsymbol{\mu}) = 0 \quad (24)$$

$$\frac{1}{n} \sum_{i=1}^n pu(d_i)(\mathbf{r}_i - \boldsymbol{\mu})(\mathbf{r}_i - \boldsymbol{\mu})' - v(d_i)\boldsymbol{\Sigma} = 0 \quad (25)$$

dengan  $d_i = (\mathbf{r}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{r}_i - \boldsymbol{\mu})$ ,

$\psi(d_i) = \frac{\partial \rho}{\partial d}$ ,  $u(d_i) = \psi(d_i)/d_i$ , sedangkan

$v(d_i) = \psi(d_i)d_i - \rho(d_i) + b_0$ . Estimasi S ditentukan oleh pemilihan fungsi kerugian (*loss function*).

Perhitungan estimasi S dilakukan secara iteratif dengan menggunakan persamaan (24) dan (25). Menurut Hardin (2000) algoritma perhitungan estimasi S adalah sebagai berikut:

1. Tentukan estimasi awal vektor mean dan matriks kovariansi,  $\hat{\boldsymbol{\mu}}_0$  dan  $\hat{\boldsymbol{\Sigma}}_0$
2. Hitung  $d_i = (\mathbf{r}_i - \hat{\boldsymbol{\mu}}_0)'\hat{\boldsymbol{\Sigma}}_0^{-1}(\mathbf{r}_i - \hat{\boldsymbol{\mu}}_0)$
3. Tentukan  $k_0$  sedemikian hingga  $\frac{\sum \rho(d_i/k_0)}{n} = b_0$
4. Hitung  $\tilde{d}_i = \frac{d_i}{k_0}$
5. Tentukan  $\hat{\boldsymbol{\mu}} = \frac{\sum \psi(\tilde{d}_i)\mathbf{r}_i}{\sum \psi(\tilde{d}_i)}$  dan  $\hat{\boldsymbol{\Sigma}} = \frac{p \sum \psi(\tilde{d}_i)(\mathbf{r}_i - \boldsymbol{\mu})(\mathbf{r}_i - \boldsymbol{\mu})'}{\sum \psi(\tilde{d}_i)}$
6. Ulangi langkah 2-3 sampai  $\hat{\boldsymbol{\mu}}$  dan  $\hat{\boldsymbol{\Sigma}}$  konvergen

## HASIL DAN PEMBAHASAN

### Hasil Klustering

Pada penelitian ini, kami menggunakan harga penutupan harian saham harian dari seluruh saham yang

termasuk dalam indeks LQ-45 yang terdaftar di Bursa Efek Indonesia periode Agustus 2017 – Februari 2018 yang kemudian diperluas hingga Juli 2018 yang diakses melalui website <https://finance.yahoo.com>. Analisis kluster yang digunakan adalah analisis kluster *time series* PAM dengan ukuran disimilaritas jarak DTW. Dengan menggunakan fungsi **TSclust** pada *packages* R, diperoleh bahwa saham-saham LQ-45 dapat dikelompokkan menjadi 6 kluster, sebagaimana disajikan pada Tabel 1.

Dari Tabel 1 dapat dilihat bahwa dengan menggunakan klustering *time series* PAM dengan ukuran disimilaritas DTW, saham-saham LQ-45 tersebar ke dalam enam kluster dengan jumlah anggota masing-masing kluster cukup variatif. Kluster 1 terdiri dari 10 saham, kluster 2 terdiri 19 saham, kluster 3 terdiri dari 11 saham, kluster 4 terdiri dari 3 saham, sedangkan kluster 5 dan kluster 6 masing-masing beranggotakan 1 saham.

### Saham Representasi Kluster

Setelah kluster-kluster terbentuk langkah selanjutnya adalah menghitung Sharpe *ratio* masing-masing saham pada setiap kluster. Dalam perhitungan Sharpe *ratio* ini, *return risk-free rate* yang digunakan adalah Bank Indonesia rate pada waktu pengambilan data, yaitu 5.25 % per tahun. Berdasarkan perhitungan Sharpe *ratio* setiap saham pada masing-masing kluster diperoleh saham-saham yang merupakan representasi masing-masing kluster untuk menyusun portofolio optimum sebagaimana disajikan pada Tabel 2.

Tabel 1. Kluster Saham

Kluster	Anggota Kluster									
1	AALI	AKRA	ASII	BBNI	BMRI	ICBP	INDF	JSMR	LPPF	SMC
2	ADHI	ADRO	BBRI	BBTN	BJBR	BRPT	BSDE	EXCL	HMSP	INC
	KLBF	PGAS	PTBA	PTPP	SCMA	SSMS	TLKM	WIKA	WSKT	
3	ANTM	BMTR	BUMI	LPKR	LSIP	MNCN	MYRX	PPRO	PWON	SMR
	SRIL									
4	BBCA	INTP	UNTR							
5	GGRM									
6	UNVR									

Tabel 2. Saham-saham Representasi Klaster

Klaster	Representasi	Return	Risiko	Sharpe Ratio
1	ICBP	0,00028	0,01395	0,00958
2	INCO	0,00258	0,02673	0,09116
3	ANTM	0,00138	0,02431	0,05068
4	BBCA	0,00088	0,01285	0,05713
5	GGRM	0,00032	0,01830	0,00941
6	UNVR	-0,00034	0,01383	-0,03521

Pada klaster 1 yang terdiri 10 saham dan saham ICBP dengan Sharpe *ratio* sebesar 0,00958 merupakan saham dengan Sharpe *ratio* tertinggi di klaster 1, sehingga saham ICBP terpilih sebagai representasi dari klaster 1. Selanjutnya pada klaster 2 yang terdiri dari 19 saham dan saham INCO dengan Sharpe *ratio* sebesar 0,09116 merupakan saham dengan Sharpe *ratio* tertinggi di klaster 2, sehingga saham INCO terpilih sebagai representasi dari klaster 2. Demikian seterusnya dengan cara yang sama, dipilih saham-saham ANTM, BBCA, GGRM, dan UNVR secara berturut-turut sebagai representasi dari klaster 3, 4, 5, dan 6.

### Perbandingan Kinerja Portofolio yang Terbentuk

Pada penelitian ini portofolio optimum ditentukan dengan menggunakan tiga model portofolio, yaitu: model portofolio MV klasik ( $MV_{klasik}$ ), model portofolio MV *robust* FMCD ( $MV_{FMCD}$ ) dan model portofolio MV *robust* S ( $MV_S$ ). Langkah pertama yang dilakukan adalah

menentukan bobot portofolio ketiga model untuk berbagai nilai *risk aversion*  $\gamma$  menggunakan fungsi **CovMcd** dan **CovSest** pada *packages* R. Saham-saham yang digunakan adalah saham-saham yang merupakan representasi setiap klaster sebagaimana disajikan pada Tabel 1. Bobot portofolio yang dihasilkan diberikan pada Tabel 3.

Dari Tabel 3 dapat dilihat bahwa, untuk model portofolio MV klasik, saham dengan return negatif yaitu saham UNVR memiliki bobot negatif (*short selling*) untuk semua nilai *risk aversion*  $\gamma$ . Sebaliknya saham-saham yang lain selalu memiliki bobot positif, kecuali saham ICBP untuk  $\gamma = 0,5$  dan  $\gamma = 1$ .

Dari Tabel 3, dapat dilihat pula bahwa untuk model portofolio MV *robust* FMCD, saham ICBP, INCO dan UNVR memiliki bobot negatif untuk hampir semua nilai *risk aversion*  $\gamma$ . Namun demikian bobot negatif tersebut semakin mengecil seiring dengan bertambahnya nilai *risk aversion*  $\gamma$ . Di sisi lain, saham ANTM, BBCA dan GGRM selalu memiliki bobot positif. Dengan

Tabel 3. Bobot Portofolio

Model	$\gamma$	ICBP	INCO	ANTM	BBCA	GGRM	UNVR
$MV_{klasik}$	0,5	-0,50075	6,01759	0,01913	7,49371	0,35268	-12,38236
	1	-0,11771	3,05234	0,02227	3,90435	0,22244	-6,08368
	2	0,07380	1,56971	0,02384	2,10966	0,15733	-2,93434
	5	0,18871	0,68014	0,02478	1,03285	0,11826	-1,04474
	10	0,22702	0,38361	0,02509	0,67392	0,10523	-0,41487
$MV_{FMCD}$	0,5	-11,64984	-2,47218	3,38129	15,32727	0,41117	-3,99770
	1	-5,72835	-1,21254	1,70727	7,80734	0,24466	-1,81838
	2	-2,76760	-0,58271	0,87026	4,04737	0,16140	-0,72872
	5	-0,99116	-0,20482	0,36806	1,79139	0,11145	-0,07492
	10	-0,39901	-0,07886	0,20066	1,03940	0,09480	0,14301
$MV_S$	0,5	-7,93013	0,21814	1,74590	11,37744	-0,87301	-3,53834
	1	-3,84454	0,13009	0,89565	5,83364	-0,40021	-1,61463
	2	-1,80174	0,08607	0,47052	3,06174	-0,16381	-0,65277
	5	-0,57606	0,05965	0,21545	1,39860	-0,02197	-0,07566
	10	-0,16751	0,05085	0,13042	0,84421	0,02531	0,11671

Tabel 4. Return, Risiko dan *Sharpe Ratio* Portofolio

$\gamma$	Return			Risiko			Sharpe Ratio		
	$MV_{Klasik}$	$MV_{FMCD}$	$MV_S$	$MV_{Klasik}$	$MV_{FMCD}$	$MV_S$	$MV_{Klasik}$	$MV_{FMCD}$	$MV_S$
0,5	0,02636	0,01959	0,01090	0,05167	0,04043	0,02222	0,11531	0,09669	0,07218
1	0,01346	0,00949	0,00536	0,01298	0,01015	0,00560	0,11687	0,09279	0,06976
2	0,00701	0,00444	0,00260	0,00331	0,00258	0,00145	0,11936	0,08470	0,06446
5	0,00314	0,00142	0,00093	0,00060	0,00046	0,00028	0,12213	0,05945	0,04695
10	0,00186	0,00041	0,00038	0,00022	0,00016	0,00012	0,11622	0,02111	0,02182

membesarnya nilai *risk aversion*  $\gamma$  bobot positif tersebut akan semakin menurun untuk mengimbangi saham-saham dengan bobot negatif.

Untuk model portofolio MV *robust* S, saham ICBP, GGRM dan UNVR memiliki bobot negatif untuk hampir semua nilai  $\gamma$ . Namun demikian bobot negatif tersebut semakin mengecil seiring dengan bertambahnya nilai *risk aversion*  $\gamma$ . Di sisi lain, saham INCO, ANTM dan BBKA selalu memiliki bobot positif semua nilai *risk aversion*  $\gamma$ . Sama dengan dua model yang lain, dengan membesarnya nilai *risk aversion*  $\gamma$ , bobot positif tersebut akan semakin menurun untuk mengimbangi saham-saham dengan bobot negatif.

Berdasarkan bobot portofolio, vektor mean dan matriks kovariansi, selanjutnya dapat ditentukan return, risiko dan Sharpe *ratio* ketiga portofolio yang dibentuk yang hasilnya sebagaimana disajikan pada Tabel 4.

Mengukur kinerja portofolio tidak bisa hanya dilihat dari returnnya saja tetapi juga harus memperhatikan risiko yang akan ditanggung *investor*. Ada beberapa ukuran yang dapat digunakan untuk mengukur kinerja portofolio, salah satunya adalah dengan menggunakan *Sharpe ratio*. Tabel 4 memperlihatkan return, risiko dan Sharpe *ratio* portofolio yang dibentuk dengan menggunakan klastering *time series* PAM dengan ukuran jarak DTW dikombinasikan dengan model portofolio MV klasik, model  $MV_{FMCD}$  dan model  $MV_S$ . Dari Tabel 4, secara umum dapat dilihat bahwa kinerja portofolio yang dibentuk dengan menggunakan saham representasi hasil klastering menggunakan PAM dengan ukuran DTW yang dikombinasikan dengan model portofolio MV klasik mengungguli

kinerja portofolio yang dibentuk dengan kombinasi model portofolio yang lain.

## KESIMPULAN DAN SARAN

### Kesimpulan

Paper ini membahas bagaimana teknik klastering dapat digunakan dalam manajemen portofolio saham dan membangun sistem untuk mendapatkan portofolio optimum. Hal ini dapat mengurangi waktu pemilihan saham karena saham dengan kategori serupa dapat dengan mudah dikelompokkan ke dalam satu klaster. Setelah klaster-klaster terbentuk, kemudian dipilih saham berkinerja terbaik dari masing-masing klaster sebagai representasi klaster untuk membentuk portofolio optimum. Studi empiris menunjukkan bahwa dengan menggunakan klastering *time series* PAM dengan ukuran jarak DTW, 45 saham Bursa Efek Indonesia yang termasuk dalam indeks LQ-45 dapat dikelompokkan menjadi 6 klaster. Saham representasi masing-masing klaster kemudian digunakan untuk membentuk portofolio optimum menggunakan model portofolio MV klasik, model portofolio MV *robust* FMCD dan model portofolio MV *robust* S. Hasil penelitian ini menunjukkan bahwa kinerja portofolio yang dihasilkan dengan menggunakan klastering *time series* PAM dengan ukuran jarak DTW yang dikombinasikan dengan model portofolio MV klasik mengungguli kinerja portofolio yang dihasilkan kombinasi dengan model yang lain.

### Saran

Untuk penelitian selanjutnya, akan menarik untuk menggunakan teknik klastering *time series* dengan ukuran disimilaritas yang lain seperti *periodogram based distance*, *Piccolo distance*, *Maharaj*

*distance* dan mengkombinasikannya dengan model-model portofolio yang sudah ada, kemudian membandingkan kinerja portofolio yang dihasilkan.

## UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih atas dukungan finansial dari Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia melalui Hibah Penelitian Disertasi Doktor (PDD) Tahun 2021. Penulis pertama juga mengucapkan terima kasih kepada Lembaga Pengelola Dana Pendidikan (LPDP) Kementerian Keuangan Republik Indonesia yang telah memberikan beasiswa untuk Program Doktor di Departemen Matematika Universitas Gadjah Mada.

## DAFTAR PUSTAKA

- Alonso, A. M., Barrendero, J. R., Hernandez, A., dan Justel, A. 2006. Time Series Clustering Based on Forecast Densities. *Computational Statistics & Data Analysis*, Vol. 51, 762–776.
- Best, Michael J. dan Grauer, Robert R. 1991. On the Sensitivity of Mean-Variance Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. *Review of Financial Studies*, Vol. 4(2), 315–342.
- Berndt, Donald J. dan Clifford, James. 1994. Using Dynamic Time Warping to Find Patterns in Time Series. In *Workshop on Knowledge Discovery in Databases*, 359–370.
- Ceria, Sabastian dan Stubbs, Robert A. 2006. Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management*, Vol. 7(2), 109–127.
- Chen, Liang H. dan Huang, Lindsay. 2009. Portfolio optimization of equity mutual funds with fuzzy return rates and risks. *Expert Systems with Applications*, Vol. 36, 3720–3727.
- Chopra, Vijay K. dan Ziemba, William T. 1993. The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. *Journal of Portfolio Management*, Vol. 19(2), 6–11.
- Davies, P. L. 1987. Asymptotic Behaviour of S-Estimates of Multivariate Location Parameters and Dispersion Matrices. *The Annals of Statistics*, Vol. 15(3), 1269–1292.
- DeMiguel, Victor dan Nogales, Francisco J. 2009. Portfolio Selection With Robust Estimation. *Operations Research*, Vol. 57(3), 560–577.
- Elton, E. J., Gruber, M. J., Brown, S. J., dan Geotzmann, W. N. 2014. *Modern Portfolio Theory and Investment Analysis*, 9th Edition, John Wiley and Sons, Inc., New York.
- Guan, He S. dan Jiang, Qing S. 2007. Cluster financial time series for portfolio. Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR 2007, Vol. 2, 851–856.
- Gubu, L. Rosadi, D., dan Abdurakhman. 2019. Classical portfolio selection with cluster analysis: Comparison between hierarchical complete linkage and Ward algorithm, in International Conference on Mathematics and its Applications 2019, *AIP Conference Proceedings* 2192, 090004-1–0960004-7.
- Hardin, Johana S. 2000. Multivariate Outlier Detection and Robust Clustering with Minimum Covariance Determinant Estimation and S-estimation. *Disertasi*, University of California, California.
- Kaufman, Leonard dan Rousseeuw, Peter J. 1990. *Finding Groups in Data: An Introduction*, John Wiley & Sons, Hoboken, New Jersey.
- Lauprete, Geoffrey J. 2001. Portfolio Risk Minimization Under Departures from Normality. *Disertasi*, Massachusetts Institute of Technology.
- Long, N. C., Wisitpongphan, N., Meesad, P., dan Unger, H. 2014. Clustering stock data for multi-objective portfolio optimization. *International Journal of Computational*



- Intelligence and Applications*, Vol. 13(2), 1–13.
- Lopuhaa, Hendrik P. 1989. On the Relation between S-Estimators and M-Estimators of Multivariate Location and Covariance. *The Annals of Statistics*, Vol. 17(4), 1662–1683.
- Markowitz, Harry. 1952. Portfolio Selection. *The Journal of Finance*, Vol. 7(1), 77–91.
- Maronna, R. A., Martin, R. D. dan Yohai, V. J. 2006. *Robust Statistics: Theory and Methods*. John Wiley & Sons, Ltd.
- Montero, Pablo dan Vilar, Jose A. 2014. TSclust: an R Package for Time Series Clustering. *Journal of Statistical Software*, Vol. 62(1), 1–43.
- Nanda, S. R., Mahanty, B., dan Tiwari, M. K. 2010. Clustering Indian Stock Market Data for Portfolio Management. *Expert Systems with Applications*, Vol. 37(12), 8793–8798.
- Reilly, Frank K. dan Brown, Keith C. 2012. *Investment Analysis and Portfolio Management. 10th Edition*, Thomson South-Western. USA
- Rousseeuw, Peter J. dan Van Driessen, Katrien. 1999. Fast Algorithm For Minimum Covariance Determinant Estimator. *Technometrics*, Vol. 41(3), 212–223.
- Rousseeuw, Peter J. dan Yohai, Victor. 1984. Robust Regression By Means of S estimators, *Lecture Notes in Statistics: Robust and Nonlinear Time Series Analysis*, Vol. 26, 256–272.
- Sankoff, David dan Kruskal, Joseph. 1983. Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison, Addison Wesley.
- Sharpe, William F. 1994. The Sharpe Ratio, *The Journal of Portfolio Management*, Vol. 21, 49–58.
- Supandi, Epha D. 2017. Pengembangan Model Portofolio Mean-Variance Melalui Metode Estimasi Robust dan Optimasi Robust. *Disertasi*, Departemen Matematika FMIPA Universitas Gadjah Mada, Yogyakarta
- Tola, V., Lillo, F., Gallegati, M., dan Mantegna, R. N. 2008. Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control*, Vol. 32(1), 235–258.
- Vaz-de Melo, Beatriz dan Camara, Ricardo P. 2005. Robust Multivariate Modeling in Finance, *International Journal of Managerial Finance*, Vol. 1(2), 95–107.
- Welsch, Roy E. dan Zhou, Xinfeng. 2007. Application of robust statistics to asset allocation models. *REVSTAT–Statistical Journal*, Vol. 5(1), 97–114.
- Winston, Wayne L. dan Goldberg, Jeffrey B. 2004. *Operations Research: Applications and Algorithms, Fourth Edition*, Thomson Learning Inc.



# SEX AND AGE GROUP DIFFERENCES IN THE SPREAD OF TUBERCULOSIS IN INDONESIA: AN AGENT BASED MODELING APPROACH

Dwi Ari Suryawan. S<sup>1</sup>, Tiodora Hadumaon Siagian<sup>2</sup>

<sup>1,2</sup> Politeknik Statistika STIS, Jakarta  
e-mail: <sup>1</sup>1211709644@stis.ac.id, <sup>2</sup>theo@stis.ac.id

## Abstrak

Tuberculosis (TB) is a serious contagious disease. Numerous research outside of Indonesia found that incidence of TB vary among different age groups and sex. This highlights the importance of research evaluating whether the spread of TB in Indonesia also differs by age group and sex. Especially considering data and information of the number and distribution of TB cases are very limited in Indonesia, as it requires certain expertise and large cost for data collection. Agent Based Modelling (ABM) is a tool that can be used to model epidemic spread. ABM has an advantage in describing structured epidemiological processes involving human behaviour and local interaction. Therefore, this study used ABM to enhance the understanding of the spread pattern of TB in Indonesia. The data used to form the model are from the publications of the Indonesia Ministry of Health, Statistics Indonesia, and WHO as from 2018 to 2020. The study results confirmed that at national level, TB incidences varied among different age groups and sexes. It is also revealed the need for further study on the spread of TB in the age group under 15 years. It is expected that these findings can be useful for TB control program in Indonesia.

**Keywords:** Tuberculosis, Agent-Based Modeling, Age Group, Sex, Indonesia

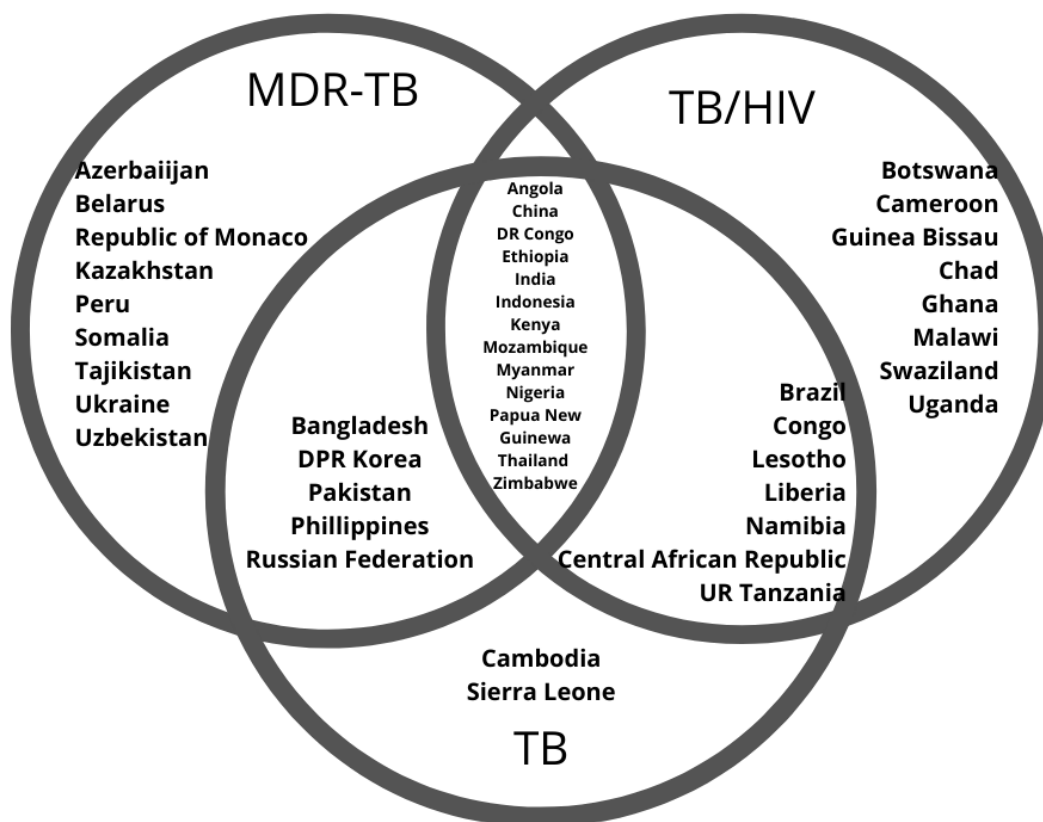
## INTRODUCTION

Tuberculosis is a disease caused by bacterium *Mycobacterium Tuberculosis* and often attacks the lungs of the sufferer and followed by a severe cough, fever, and a heavy feeling in the chest. (Fogel, 2015; WHO, 2020). Commonly, people with tuberculosis have two kinds of conditions, latent (hidden) infected conditions and active infected conditions (Institute of Medicine (US), 2001). *Tuberculosis* or TB can spread from person to person through the air, therefore TB spreads very quickly.

In 2018, It is known that a quarter of the world's population have infected with the bacteria that causes TB. (WHO, 2020). Based on the publications of the World Health Organization, TB is included in the top 10 diseases that cause death worldwide. Due to its wide spread and high number of cases, TB is classified as a pandemic (*Global Pandemic | TB Alliance, 2022*). It needs special attention for the entire

population of the earth. Therefore, the United Nations has decided on TB eradication, and targeting in 2030 as part of the Sustainable Development Goals (SDGs).

The WHO report in 2019 shows that Indonesia is included in the 14 countries with a high burden of TB, together with China, India, Nigeria and other developing countries, especially for the Southeast Asia region (Figure 1). In 2018, it is estimated that there are 842,000 TB cases and around 32% of those cases have not been reported in Indonesia (Kemenkes RI, 2018). Indonesia also included in the group of countries with high cases of Multi-drug-resistant TB (bacterial resistance to at least two types of anti-TB drugs). In addition, Indonesia also has high active TB and HIV positive cases at the same time. With the high cases of TB, MDR-TB, and being positive for TB and HIV simultaneously, Indonesia is classified as one of the 30 countries with a heavy TB burden (WHO,



source: WHO, 2020

Figure 1. High Burden Countries by TB, TB/HIV, and MDR-TB According to WHO 2016-2020

2019). This indicates that serious action is needed for Indonesia to immediately deal with TB.

TB can affect anyone, regardless of ethnicity, gender, and social status (Rajagopalan, 2001). However, several studies have shown that cases of morbidity and mortality due to tuberculosis are more vulnerable in the elderly population (Negin et al., 2015; Rajagopalan, 2001; Schaaf et al., 2010). In addition, the data show that more male are sick with TB (WHO, 2019; Horton et al., 2016).

The number of cases of TB disease among elderly population is due to the difficulty of detecting TB. This is due to the weak immune system of the elderly population (Schaaf et al., 2010; Negin, 2015). Data from the BPS shows that in 2020 in Indonesia around 14.8% of the population are people with an age of more than 55 years. In addition, based on WHO data in its publications, it shows that more adults and the elderly are exposed to TB. This indicates that special attention is needed for the elderly population to avoid and treat TB (WHO, 2020).

Several studies and publications also show that gender has an effect on the incidence of TB (WHO, 2020; Horton et al., 2016; Mangngi, 2019). The difficulty of detecting TB in the elderly population and the large number of TB cases in the male population, are a challenge in seeing and determining the distribution of TB cases in the elderly population and the differences by sex. The ability to collect TB data through survey is very limited, as more capital and specific expertise are needed than usual survey. In addition, collecting TB data using surveys is difficult, especially in Indonesia. One of the reasons is that there are many symptoms of diseases in the elderly population may lead to a misdiagnosis. A method that can be used to model TB cases is Agent Based Modelling (hereinafter is called ABM). Pappalardo et al. in 2019, for instance, they successfully made an extension for ABM named Universal Immune System Simulator (UISS) which is applied to large number of disease cases to analysis tuberculosis –

immune system dynamics. In Indonesia, there have been studies that used ABM for TB cases too, such as that conducted by Gema Wahyu Saputra, Budhi Irawan, and Purba Daru Kusuma in 2018. In addition, there were also studies using ABM in a structured age population conducted by Graciani Rodrigues, Espíndola and Penna, in 2015. Agent-based simulations can be used to obtain information about TB cases in the elderly population based on gender in Indonesia, such as for estimating TB cases, viewing a graph of the development of TB cases, or even comparing TB cases in the elderly and non-elderly population age. Agent-based simulation or also known as Agent Based Modeling (ABM) is a method that allows researchers to create, analyze, and conduct experiments from a model consisting of a number of agents. (Siagian and Prasojo, 2021). By simulating, some information or data can be retrieved without conducting a more expensive survey or census. The results of the simulation are expected to provide an overview of the distribution of TB in the population that classified to the age group and the differences by gender. Taking this into account, further research on the distribution of TB by sex can be carried out using the ABM method. To date there are no studies related to TB that classified to the age group in Indonesia and the differences by gender using the ABM method. Therefore TB research by age group difference and the differences according to sex in Indonesia is needed with the help of the ABM method. This research may help to analyze how the condition of TB in Indonesia, so the government and society can aware and act correctly to reduce the TB cases in Indonesia.

With the existing problems, there were several questions related to the ABM research and its application in cases of TB Indonesia and the difference between ages and the sexes. Some of the questions are as follows: 1) How is the ABM model for the spread of TB cases in Indonesia; 2) How is the description of TB cases among the sick, under treatment and cured population in Indonesia; 3) How are TB cases different

between male and female for elderly population in Indonesia; 4) Is the model formed valid or not?

To answer the existing research questions, the objectives of the research are described as follows: 1) Build an ABM model for the spread of TB in Indonesia. 2) Conducting simulations with the ABM model to get an overview of TB cases in the sick population, under treatment, and cured from TB in Indonesia. 3) To determine the pattern of TB in the age groups and the differences by gender. 4) Validate the ABM Model for TB spread in Indonesia.

## DATA AND STUDY AREA

ABM processed using Netlogo software. Netlogo is free and open source software developed by Northwestern University. In addition, the ABM model used for the formation of ABM in TB cases in the age groups and its differences according to gender in Indonesia, is a development of the contagion model.

For the validation stage of the model which is formed, it is done by comparing the results of the ABM model that is formed with the published official statistics. The output of the ABM results are in the form of graphs and statistics such as percentages and summaries. From the graphs and percentages generated through the ABM, then it compared with the published statistics.

The variables or statistics that are inputted or used before the simulation process is carried out are as follows:

1. Number for total population in the model (Indonesian population scaled).
2. Rate of TB infection per 100000 population
3. The time length of treatment which carried out.
4. Sick TB rate per 100000 population.
5. Rate under treatment per 100000 population
6. Rate treated per 100000 population
7. Simulation time (in years)
8. Aids-rate Indonesia
9. Diabetes rate
10. Factors resistant to TB drugs
11. Abandon rate

12. Diabetes-factor
13. Risk Factor
14. Mean diagnosis delay
15. Correction factor

The variables obtained during the simulation process are as follows:

1. Movement (in the form of graphs) of the addition or decrease of agents with the status of infected, sick, under treatment and cured.

The variables obtained after the simulation process ends are as follows:

1. Prediction in the form of elderly population and non-elderly population based on sex with sick status.
2. Descriptive statistics of the elderly population and non-elderly population by sex with sick status

These data were collected from the publications of the Ministry of Health, BPS, and WHO from the latest publication, it comes from 2018 to 2020. In addition, the characteristics and behavior of agents regarding TB in Indonesia are adjusted to literature studies and publications from the Indonesian Ministry of Health and BPS. Information in the form of population and population structure is used to form an artificial population that resembles the characteristics of the Indonesian. So, a simulation is created that is similar to the real condition.

The number of agents is also adjusted to the population in Indonesia. Agents will be distinguished according to the status given, namely healthy, infected, sick, under treatment and cured. The movement of the agent will be simulated in two periods, namely 360 days and 720 days.

Supporting statistics are inputted on the interface of the ABM model which is made to resemble the situation or condition in Indonesia. Other supporting variables such as diabetes rate, Aids, abandon rate and others are obtained from publications from the Ministry of Health. Some variables or statistics are used at the back of the model or coding such as variable of proportion of the population suffering from TB according to gender and age. In addition, there are also probability for women and men to be infected with TB in

Indonesia. Counting the number of days for each agent to change status is also set on the back section or coding.

After the process of forming the ABM has been completed and the required statistics have been inputted, the model will produce output in the form of statistics of agent with sick status. The statistics generated are in the form of population movements with sick status during the simulation, according to age and gender. Then, the movement will be divided into elderly and non-elderly residents, as well as men and women. The elderly in this category are residents of population aged 65 years and over.

## METHODOLOGY

ABM is a model with a bottom-up type, where the smallest units are agents who act according to the program that has been made. The characteristics of the agent are made according to the characteristics of the Indonesians. In this study, the ABM model used is an application or continuation of the ABM model on TB cases in Barcelona made by Julia Vila Guilera, Clara Prats, and Angels Orcau. Julia, et al. in his research formed the ABM model of TB epidemiology in Barcelona by considering the role of age, gender and origin of the population (Vila Guilera, 2017).

Agent-based modeling in its application is a model that uses a computer program to create a number of agents with programmed characteristics and behaviors. The purpose of ABM is to observe what will happen to the simulation progresses. The elements that make up ABM are a number of agents, rules or choices that agents will make (e.g.: move, get infected, die, etc.), interaction structures such as environment, and randomness.

The agents used in this study are imitation individuals from Indonesian population. Individuals have several attributes such as age group and gender. Each agent or individual will be given an order to interact with each other. From the interactions between agents, it can be seen how the virus spreads from agent to agent.

The basic model used to make the ABM for the spread of TB in Indonesia in the difference according age and gender is the contagion model. The contagion model is commonly used to model various problems such as the spread of disease, the spread of information, the spread of hoaxes, various health-related behaviors, and others. In this study, the contagion model was used to see the spread of TB disease in Indonesia. In addition, the type of epidemic model used in this research is SEIR (Susceptible, Exposed, Infectious, and recovered) model.

The first stage is a literature study to study the ABM and TB methods in Indonesia. To learn more about ABM, the website [ccl.northwestern.edu/](http://ccl.northwestern.edu/) provides and gives information related to ABM and the applications that support it, such as Netlogo. Furthermore, to deepen about TB in Indonesia using TB-related research and publications from related research.

The formation of the ABM model is divided into three main parts, the three main parts are: (1) *specification – formalization*, (2) *Modelling – Verification – Experimentation*, and (3) *Calibration – Validation* (Salgado and Gilbert, 2013). In the first stage is to reviews the terms of the theories that will be used to answer the research question and turn it into a mathematical or logical expression. The next step is the formation of the model with the help of the netlogo application. The final step is to test or validate the results of the model formed.

The first stage in this research is to collect related theories and studies regarding ABM and TB. The aim is to study and support the results of the research that will be obtained. In addition, from the theories and related research collected, research questions are obtained which will later be answered by establishing the ABM model. After the theories are collected, then they are changed or specified into a logical or mathematical form so that further steps can be taken.

The second stage in the formation of the ABM is modeling, verification, and experimentation. At this stage the thing to

do is to change the logical or mathematical statements that have been formed in the previous step into a computer program. In this study using a computer program or Netlogo software. Then check the program that has been made. Whether there is an error or there is a discrepancy with the existing theory, this step is called verification. This is necessary because every time you run the program, you may get different results each time you run it. The snippets of the coding used in coding section is as follows:

```

set get-sick-rate 0
set from-infected-to-healthy from-
infected-to-healthy + 1
ask patch-here [ set num-healthy num-
healthy + 1]
die
]
; Different sickening probabilities are
assigned depending on the age of the
infected turtles. 50% of the 0-5 y.o
infected will sicken, 15% for the 5-15
y.o and 3% for the 15 y.o plus.
;
if age <= (5 * 365) [set age-factor 1]
if age > (5 * 365) and age <= (15 * 365)
[set age-factor 1]
if age > (15 * 365) [set age-factor 1]
if gender = 1 [set gender-factor 1]
if gender = 0 [set gender-factor 1]

```

P.S: The code section is too long to be display, here are little bit of part that has been used.

Calibration and Validation is the last stage of the formation of the ABM model aims to test or validate the model that has been formed. According to Bianchi, Cirillo, Gallegati, & Vagliasindi, (2007) in the research of Salgado & Gilbert in 2013 stated that there are three ways to validate the ABM model, namely Descriptive output validation, Predictive output validation, and Input validation. In this study, we use descriptive output validation, which matches the output of the computational results of the ABM model with available data, in this case statistics from publications from the Indonesian Ministry of Health and WHO.

## RESULT AND DISCUSSION

### Formation of ABM Model

Theoretical models can be defined as the development of theories that aim to explain a situation or phenomenon, some of which can even do forecasting. In social science, model is a mathematical abstraction or simplification of a social process. Some of the purposes of modeling are to understand, measure, and estimate variable(s). Based on the type, models in social science are divided into three; statistical models, analytical models, and agent-based models (Balietti, 2012; Lee, 2012). One of models that can be used in social science is ABM. ABM uses computer programs to form a number of agents that have certain characteristics and behaviors, in order to observe what will happen over time. ABM has begun to be widely used in many fields, one of which is social science, such as in the case of disaster evacuation simulations, forest fires, and even the spread of diseases such as Covid-19 or TB

TB is a fairly complex epidemic disease. In its spread, TB is influenced by many factors such as demographics, health, environment and others. As an effort to understand TB cases in Indonesia, an ABM model for TB cases was created that focuses on TB cases in the Indonesian population and the differences according to gender and age groups.

In this study, the ABM model used is a development or continuation of the ABM model on TB cases in Barcelona made by Julia Vila Guilera, Clara Prats, and Angels Orcau. Julia, et al. in his research formed the ABM model of TB epidemiology in Barcelona by considering the role of age, gender and origin of the population (Vila Guilera, 2017). What distinguishes this research from (Vila Guilera, 2017) research is that this study tries to form a model with the characteristics of the Indonesian region and focuses on the elderly population and their differences by gender. This is deemed necessary considering that the formation of the ABM model for TB cases in Indonesia is still relatively rare, especially for research that focuses on the elderly population and the differences based on gender.



In general, the steps to form the ABM model are divided into three steps (Salgado and Gilbert, 2013). The first step is to collect related theories and research on TB cases, especially in Indonesia. Next is modeling, which is changing the logical statement obtained from the first step into a computer program, in this study a logical statement or logic change is made into the Netlogo programming language. The last step is model validation, the goal is to determine whether the model formed is appropriate or resembles the original state, in this study validation is carried out by comparing the output of the model that has been formed with statistics from the publication.

Theories and related research are collected to form a solid basis in the formation of the model. The result of the theory and related research collected is that in the case of TB it is influenced by many factors. Several factors influence TB and are used in the model such as diabetes, aids, immunity to TB drugs, patients who ignore treatment, the average time the patient is checked for illness, the length of time to recover from TB, and other risk factors. In addition, the number of TB infection, illness, treatment period, and recovery from TB per 100,000 population is also used, simulated population data and simulation time are also used.

Through a study of theories, related research, and publications obtained, it is known the statistical values used as the basis of the model. The statistics obtained are used in two different sections, the first

in the coding section and the second in the interface section. The proportion of the Indonesian population by age and the proportion of the population by age with TB and the proportion of the population with TB by sex are used in the coding section. Others statistics are used in the interface section according to Figure 2.

In the interface, we can input the required numbers directly. The figures required have been adjusted to the conditions in Indonesia in general. For the part of the population infected with TB, for example, a value of 25,000 per 100,000 population is entered, this is based on the WHO publication which states that a quarter of the world's population is infected with the TB virus. (WHO, 2020) Other numbers such as diabetes factor, abandon rate, immunity and other factors are numbers in percent. As for the length of treatment, the mean delay-diagnosis has a unit of time, which is counted in days.

The second stage in the formation of the ABM model is modeling, verification, and experimentation. At this stage, the transformation of logical and mathematical statements into a computer program is carried out. This research converts statements into programs that can be read by the ABM processing application, Netlogo. The statements are converted to the form of a computer program in the coding section. The basic coding used was sourced from Juvia's research for TB cases in Barcelona, which was then adapted to the situation and conditions in Indonesia. After the coding section has been successfully

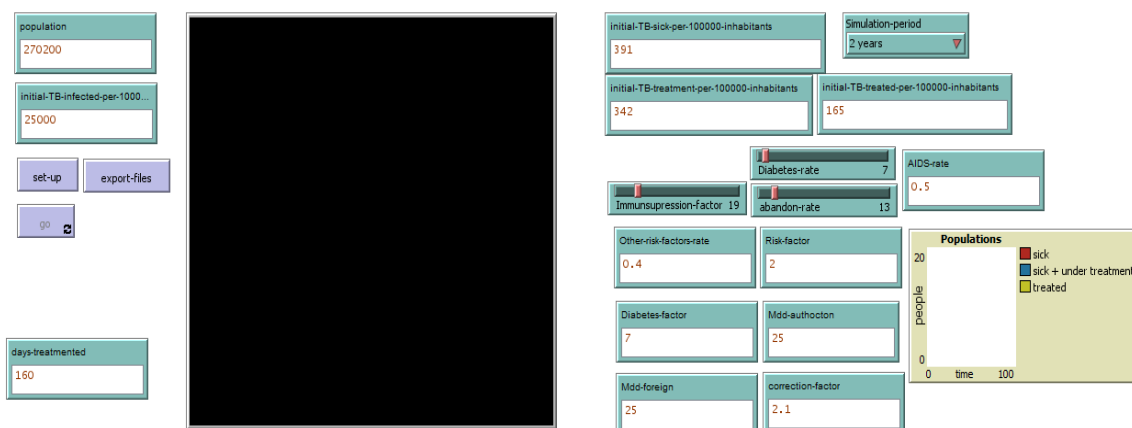


Figure 2. ABM interface section for TB cases in Indonesia

completed, then it is connected to the interface section that has been made according to Figure 2. If both the coding section and the interface have been connected, then running is done to see if there are any discrepancies or errors that occur. When no discrepancies or errors are found, the formation of the model can proceed to the next stage, namely model validation.

Model validation is a process or stage of proof of a model that can be considered appropriate (Tsiptsias, Tako and Robinson, 2016). In this study, the validation stage was carried out at the end of the research section. Validation is done by matching the results of the model that has been made with published statistic. In addition, the ABM model that is formed also has an output in the form of simulation results for cases that occur in residents with certain age groups, as well as cases that occur in female and male population. The resulting output is then matched with the latest publications published by the WHO. If the results of the output are matched with the publications, then the model that have been made can be concluded to be valid, so that it is good to use.

### Simulation with ABM Model for Agents with Sick, Under treatment, and cured Status

The simulation is carried out in two time spans, one year and two years. The purpose of the simulation is to see the movement or get an overview of TB cases with the characteristics of the population in Indonesia. To see the description and

movement of TB cases in Indonesia, a line graph with three main lines used. The line for agents with sick status (red), sick plus undertreatment (blue), and cured (yellow). The results of the simulation carried out are in accordance with the following figure.

The results of the simulations are shown in Figure 3. and it show interesting results. Based on the graph in Figure 3. it can be seen that there was a change from TB cases which initially increased to decreased in the first 25 days. This was because the ABM modeling was done setting the average duration of diagnosis for 25 days, meaning that the average duration of diagnosis was 25 days. On average, agents will perform tests or treatment after 25 days feel they have TB symptoms. This is based on the cough classification which states that an acute cough will heal in less than two weeks and the conditions or characteristics of the Indonesian population in taking treatment. (Kasi and Kamerman-Kretzmer, 2019). The reason for determining the average examination time based on the length of cough is because one of the symptoms of TB disease is chronic cough (Loddenkemper, Lipman and Zumla, 2016). In addition, from the simulation results its gained information; that the length of time the agent reports the disease or the length of time to detect a sick agent are affecting the movement of TB. This indicates that the faster it is to detect agents with TB disease, the faster the decrease in TB cases will occur. This is supported by the paper of Frank, R which states that the earlier the diagnosis is made, the higher the chances of recovering early (Frank, 2015).

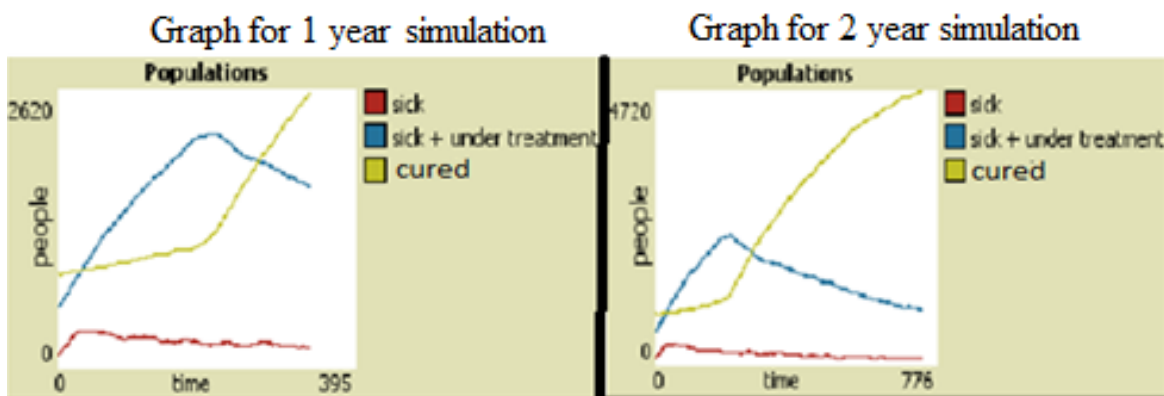


Figure 3a. ABM Simulation of TB Cases in Indonesia for 1 year (left)

Figure 3b. ABM Simulation of TB Cases in Indonesia for 2 year (Right)

Another thing that can be seen is that there is an increase in agents with under treatment status, and a decrease in cases of agents with sick status, as well as an increase in agents who have cured or been treated after the first 25 days of the simulation. The increase in the graph of agents with under treatment status is due to the change from agents with sick status to under treatment. In addition, as previously discussed, the increase in agents with under treatment status is due to the average agent inspection time. That is 25 days. The graph increase for agents with under treatment status occurred from the 25th day to the 200th day. This drastic increase in the graph indicates that facilities are needed to support the needs of patients during TB treatment during this time span. Then after day 200 there will be a decrease in the graph of agents with under treatment status followed by an increase in the graph of agents with cured status

Another finding that can be seen from the model simulation results is an increase in the graph of agents with cured status and a decrease in the graph of agents with sick and under treatment status after the 200th day. This is based on the change in status from agents with sick and under-treatment status to cured. In other words, agents who are under treatment have recovered or cured from TB disease. The increase in this graph occurs because usually individuals will fully recover from TB if the individual routinely and

disciplined follows treatment procedures during the treatment period for 160 days to 185 days or 6 to 7 months. (Kementerian Kesehatan, 2016; Terefe and Gebrewold, 2018).

The simulation is divided into three phases. The first phase occurs in the first 25 days, in this phase there is an increase in the agent with sick status. Furthermore, in the second phase what happened was a decrease in agent with sick status followed by an increase in agent with under treatment status, in this phase it occurred after the 25th day to the 200th day. The third phase is the phase where there is a decrease in agents with under treatment and sick status, on the contrary there is an increase in agents with cured status, this phase occurs after the 200th day, indicating that many individuals have successfully recovered from TB disease.

#### The Pattern of TB Distribution by Age Groups and the Differences by Gender

The pattern of TB distribution is known to vary by age group. It is known that the adult population is more exposed to cases of illness due to TB. This is in line with the results of the ABM simulation, with the following results below.

The results of the simulation for TB cases in Indonesia as shown in Figure 4. It shows the differences between TB cases according to age

Groups. The highest cases occur in the population aged 5-14 years, this could

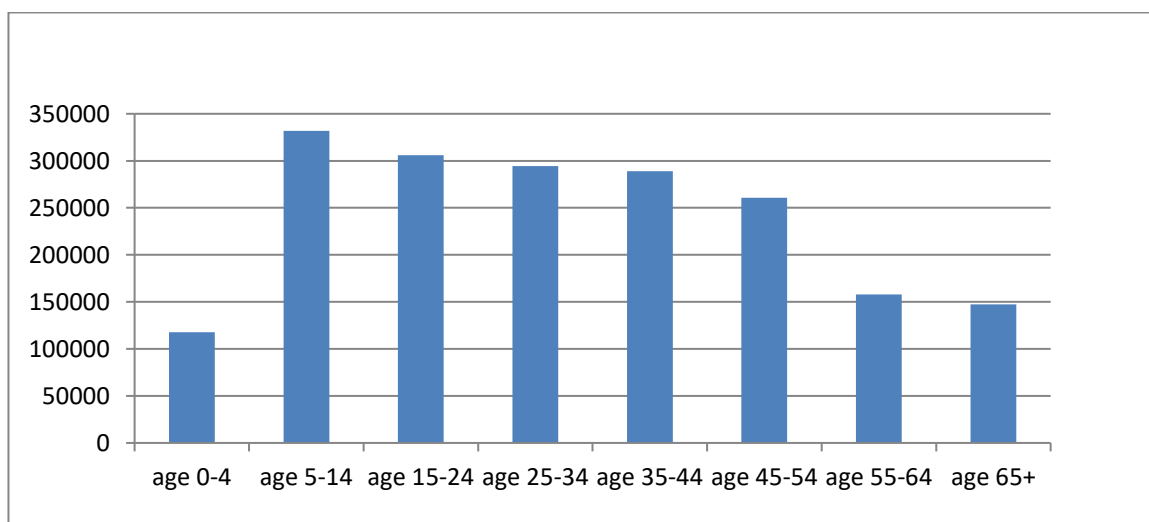


Figure 4. TB Cases by Age Group From Results of Simulation

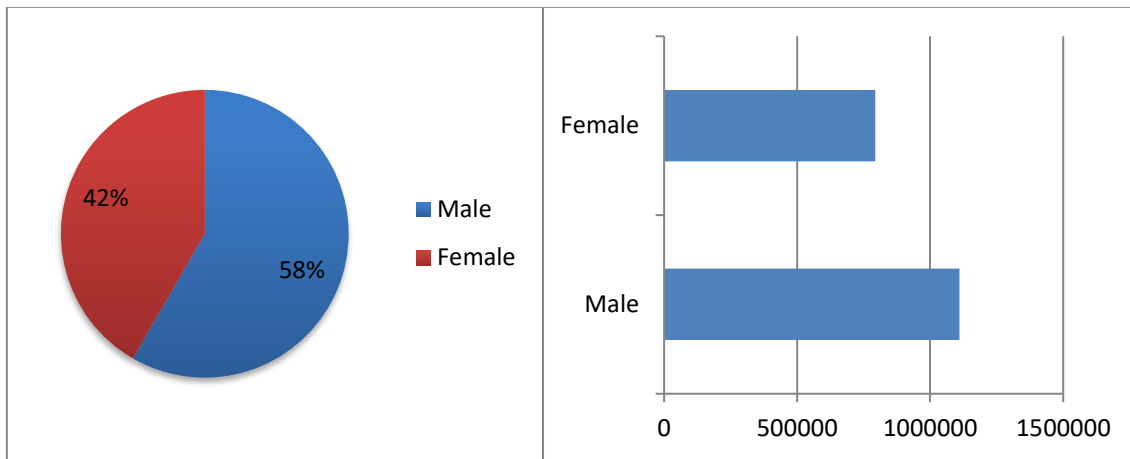


Figure 5a. Agent With Sick Status by Gender (Percentage) (Left)

Figure 5b. Agent with Sick Status by Gender (count) (Right)

be because the population under the age of 20 years has a greater chance of getting sick when infected (Vila Guilera, 2017). In addition, the population aged over 65 years does not appear to be too high compared to other age groups, but still needs to be considered, because mortality is more vulnerable in the elderly population (Negin, Abimbola and Marais, 2015). The productive age group when combined, has a large number of cases, this could be due to the large number of contacts made to other people in public places such as offices, etc. (Tuite *et al.*, 2017). Another thing besides the different age groups in the TB distribution pattern is gender. More TB cases occur in men than women (Horton *et al.*, 2016). This is in line with the results of

the ABM simulation as shown in the following graphs and charts.

The results of the ABM simulation in Fig. 5a and Fig. 5b shows that cases of TB disease that occur in males are higher than females. In addition, the simulation of the ABM that was carried out showed that men had about 16% of cases of illness due to TB compared to women. This is in line with Mangngi's research which states that men have a greater tendency to get TB disease than women (Mangngi, 2019). The results of the ABM simulation also show that differences in TB cases by gender occur in all age groups, which is visualized in the following figure.

The ABM modeling has shown that agents with male attributes get more TB

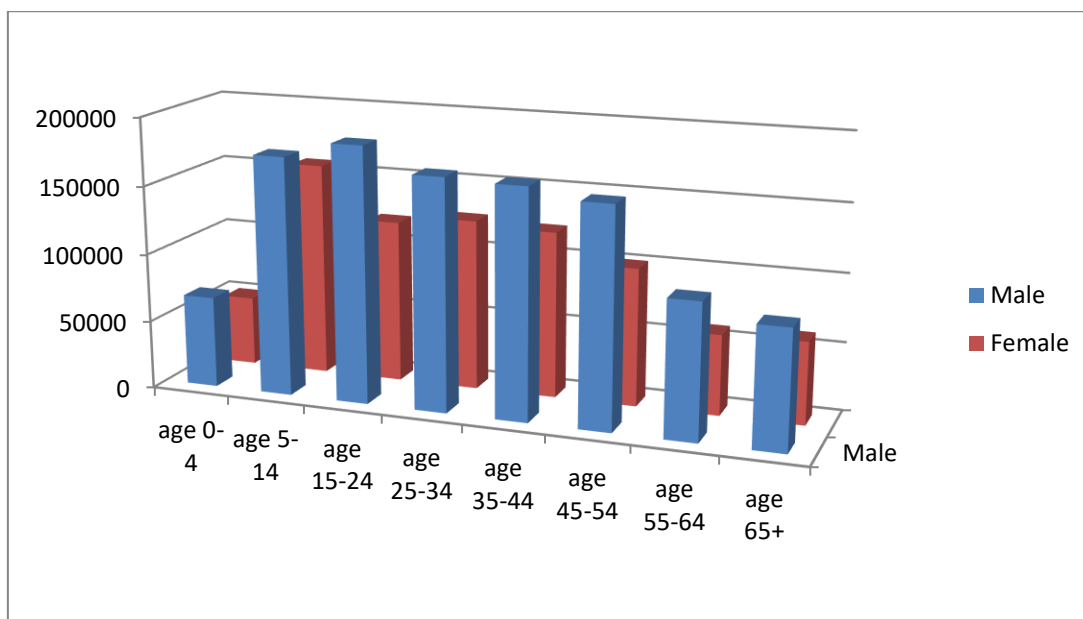


Figure 6. TB Disease Cases by Age Group and Gender

cases in all age groups. Men tend to have a lifestyle that makes them more susceptible to exposure to TB. In addition, men's lungs tend to be larger than women's, so the possibility of harboring TB-causing bacteria is higher. (Yates and Atkinson, 2017). More attention needs to be given to TB that occurs in men and people in the age group over 65 years. Although graphically the cases that occur in this age group are not too high compared to cases in other age groups, cases of TB disease in the elderly population, especially men, remain a frightening specter considering the resilience of the population of this age group is not too strong and the style of The life of men tends to make health vulnerable, followed by disease complications that may occur. Comparison of TB disease cases in the population aged over 65 and the differences by sex can be seen in more detail in the following figure.

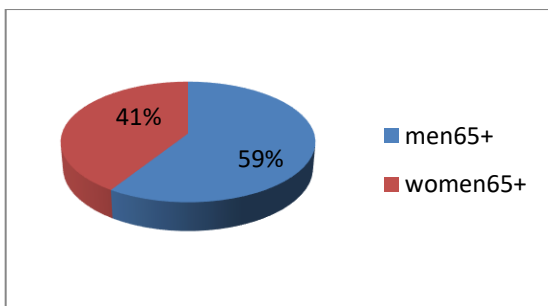


Figure 7. Percentage of Population Sick of TB in the Elderly by Gender

The graph from the ABM simulation shows that in the elderly population the percentage of TB cases is much more experienced by men. Overall it is almost the same as TB cases in general, only in the elderly the percentage of TB disease for men is greater, namely 59% compared to

58% overall. Some studies say this could be influenced by hormones and immunity possessed by men, especially those who are over 65 years old. (Neyrolles and Quintanamarci, 2009; Negin, Abimbola and Marais, 2015).

### Model Validation

Validation is carried out to determine whether the ABM model formed is a good model, so that the results of the simulation can be trusted. In this ABM model for TB cases in the elderly and the differences by gender, validation was carried out by matching the results from simulations with publications. The results that will be matched to the ABM model formed are the percentage of the elderly population compared to the non-elderly population who are sick with TB.

Broadly speaking, the results from the simulation and publications are different, although the difference is not too far, this result is shown in Figure 8. The simulation shows that the elderly population who is sick with TB is around 8% of the total. Meanwhile, the results of WHO publications show that in 2019 around 11% of the elderly population was sick with TB. When compared to the published results with simulations for TB sick cases in the elderly population, there is a difference of 3%. According to the researcher, this figure is not too large, considering that in general, for inferential analysis, an error significance level of 5% is used.

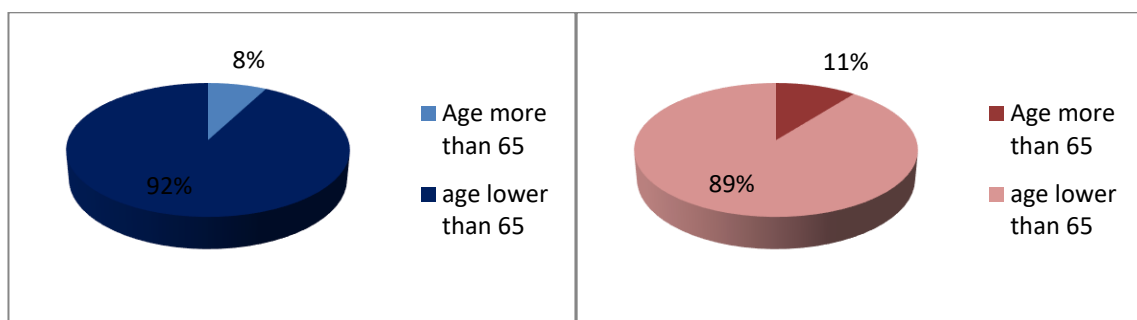


Figure 8a (Left). Comparison of the Percentage of TB Sick Cases by Age Group (Simulation)  
 Figure 8b (Right). Comparison of the Percentage of TB Sick Cases by Age Group (Publication)

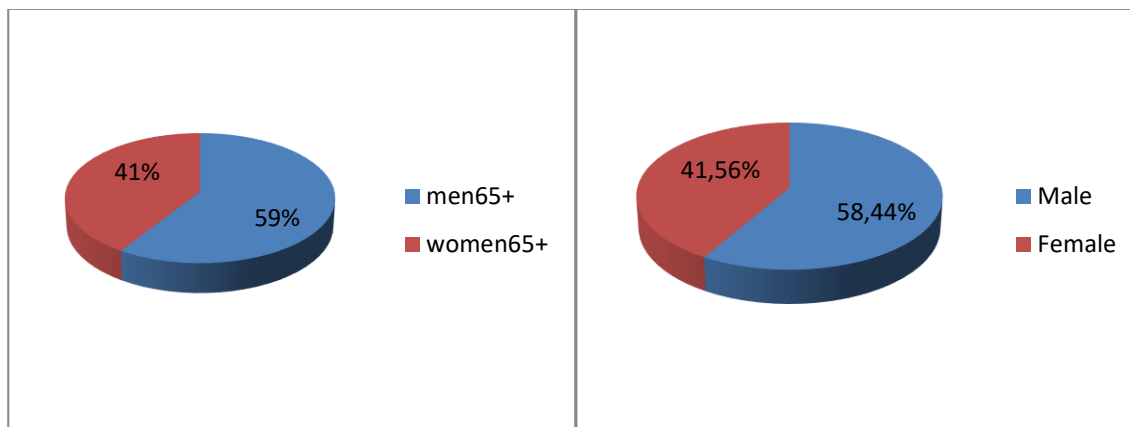


Figure 9a (Left). Pie Chart Percentage of Illness Due to TB in the Elderly by Gender.(Simulation)

Figure 9b (Right). Percentage of Illness Due to TB in the Elderly by Gender.(Simulation) (Publication)

Validation was also carried out on the results of TB cases according to gender, in addition to age group. Validation is carried out in the form of matching the percentage of simulation results and publications, the same as before. From the results of matching simulations and publications that have been made, the following results are obtained.

The results of simulations and publications for sick TB cases are shown in Figure 9a and Figure 9b. show very similar or close to the same results. Based on the results of the publication, the male population with TB in 2019 was 58.44%, while the results from the simulation showed a result of 59%. With a difference of less than 5%, it can be said that the results of the ABM model are close and reasonable when compared to the results from publications.

In general, although not exactly the same, the results of the ABM simulations have similar characteristics or similarities with the original results, when compared to publications conducted by WHO. For cases of TB disease in the elderly population, the ABM simulation results show lower results than the published results. Meanwhile, the results of the ABM simulation on the population by gender gave very similar results to the results of WHO publications, with a difference of even less than 1%. With these results, it can be concluded that the ABM model formed is a valid model, so the results can be used.

It should be admitted, this research still has many limitations. In this research, model simulations are run for national coverage, whereas, it is well understood that the characteristics between regions in Indonesia, such as between provinces, between districts/cities are very different. This is one of the limitations of this research. It is better in the future, the ABM model that is formed to be more specific for each region that wants to be studied more deeply about TB cases. In addition, this research was conducted only by visually descriptive analysis and focused on the elderly population, even though there are still many things that can be done with the formation of the ABM model that has been made. The statistical sources used in the formation of this model come from various sources such as WHO, BPS, and the Indonesian Ministry of Health because the sources obtained are different, this is a limitation of the research. Some other shortcomings are the use of statistics on the part of the population, in the ABM model that is formed the population used is a reduced-scale population statistic, this is a shortcoming considering that the more appropriate the model is with its original state, the better the model is. The validation part is carried out only by comparing the similarities visually and not statistically, so that in terms of conclusions it is still not good enough. These are the limitation of the research, but at least this research can be



used as an initial model to be continued or refined in the further research.

## CONCLUSION AND SUGGESTIONS

The ABM model for TB cases in Indonesia was successfully formed with the help of the netlogo application. The simulation results show that TB cases are divided into three phases. From the simulation results, it can also be seen that for the elderly population, TB cases are not too high compared to other age groups. In addition, the results of the simulation show that male agent are more exposed to TB disease more than female agents. The validation of the ABM model that has been carried out shows that there is a similarity between the simulation results and the publications data, which means the model can be used.

Some suggestions that can be obtained from the research results are as follows. For the government, to keep focus on eliminating to TB cases in Indonesia, especially for the elderly population and the male population. This is considering the results of the ABM simulation which shows that men are more susceptible to TB disease. So, it is important for men to be more aware of TB because TB can be very dangerous for the body if not treated immediately. As the simulation results show the largest cases among population aged less than 15 years. Further research can focus on TB cases among population with the age group less than 15 years, further research can focus on TB cases among population aged less than 15 years. In addition, further ABM modeling on TB cases that can be carried out is by using the latest statistics and be more specific at province level or regional.

## REFERENCES

- Baliotti, S. (2012) *Agent-Based Modeling*. doi: 10.1007/978-3-642-24004-1.
- Fogel, N. (2015) 'Tuberculosis: A disease without boundaries', *Tuberculosis*, 95(5), pp. 527–531. doi: 10.1016/j.tube.2015.05.017.
- Frank, R. (2015) 'IMPORTANCE OF EARLY DIAGNOSIS AND TREATMENT OF LARYNGEAL TUBERCULOSIS', pp. 3–6.
- Graciani Rodrigues, C. C., Espíndola, A. L. and Penna, T. J. P. (2015) 'An agent-based computational model for tuberculosis spreading on age-structured populations', *Physica A: Statistical Mechanics and its Applications*, 428, pp. 52–59. doi: 10.1016/j.physa.2015.02.027.
- Horton, K. C. *et al.* (2016) 'Sex Differences in Tuberculosis Burden and Notifications in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis', *PLoS Medicine*, 13(9), pp. 1–23. doi: 10.1371/journal.pmed.1002119.
- Institute of Medicine (US) (2001) *Tuberculosis in the Workplace*, National Academies Press (US).
- Kasi, A. S. and Kamerman-Kretzmer, R. J. (2019) 'Cough', *Pediatrics in Review*, 40(4), pp. 157–167. doi: 10.1542/pir.2018-0116.
- Kemenkes RI (2018) 'Tuberkulosis ( TB )', *Tuberkulosis*, 1(april), p. 2018. Available at: [www.kemendes.go.id](http://www.kemendes.go.id).
- Kementerian Kesehatan (2016) 'Info data dan informasi Tuberculosis 2016', p. 12. Available at: <http://www.pusdatin.kemkes.go.id/foilder/view/01/structure-publikasi-pusdatin-info-datin.html>.
- Lee, I. (2012) 'Agent Based Modeling - YouTube'. Available at: [https://www.youtube.com/watch?v=XB4e-3nws1g&ab\\_channel=MartinHilbert](https://www.youtube.com/watch?v=XB4e-3nws1g&ab_channel=MartinHilbert).
- Loddenkemper, R., Lipman, M. and Zumla, A. (2016) 'Clinical aspects of adult tuberculosis', *Cold Spring Harbor Perspectives in Medicine*, 6(1), pp. 1–26. doi: 10.1101/cshperspect.a017848.
- Mangngi, M. P. (2019) 'Faktor Risiko Umur, Jenis Kelamin Dan Kepadatan Hunian Terhadap Kejadian TB Paru Di Puskesmas Naibonat Tahun 2018'. Available at: <http://repository.poltekeskupang.ac.id/1936/>.

- Negin, J., Abimbola, S. and Marais, B. J. (2015) 'Tuberculosis among older adults - time to take notice', *International Journal of Infectious Diseases*, 32, pp. 135–137. doi: 10.1016/j.ijid.2014.11.018.
- Neyrolles, O. and Quintana-murci, L. (2009) 'Sexual Inequality in Tuberculosis', 6(12). doi: 10.1371/journal.pmed.1000199.
- Pappalardo, F. *et al.* (2019) 'An agent based modeling approach for the analysis of tuberculosis - Immune system dynamics', *Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018*, pp. 1386–1392. doi: 10.1109/BIBM.2018.8621355.
- Rajagopalan, S. (2001) 'Tuberculosis and aging: A global health problem', *Clinical Infectious Diseases*, 33(7), pp. 1034–1039. doi: 10.1086/322671.
- Salgado, M. and Gilbert, N. (2013) 'Agent based modelling', *Handbook of Quantitative Methods for Educational Research*, pp. 247–265. doi: 10.1007/978-94-6209-404-8.
- Schaaf, H. S. *et al.* (2010) 'Tuberculosis at extremes of age', *Respirology*, 15(5), pp. 747–763. doi: 10.1111/j.1440-1843.2010.01784.x.
- Siagian, T. H. and Prasojo, A. P. S. (2021) 'Agent-Based Modelling Pada Studi Kependudukan: Potensi Dan Tantangan', *Seminar Nasional Official Statistics*, 2020(1), pp. 1032–1040. doi: 10.34123/semnasoffstat.v2020i1.591
- Tb Alliance (no date) *Global Pandemic | TB Alliance*. Available at: <https://www.tballiance.org/why-new-tb-drugs/global-pandemic> (Accessed: 27 January 2022).
- Terefe, A. N. and Gebrewold, L. A. (2018) 'Modeling Time to Recovery of Adult Tuberculosis (Tb) Patients in MizanTepi University Teaching Hospital, South-West Ethiopia', *Mycobacterial Diseases*, 08(01), pp. 1–6. doi: 10.4172/2161-1068.1000258.
- Tsioptsias, N., Tako, A. and Robinson, S. (2016) 'Model validation and testing in simulation: A literature review', *OpenAccess Series in Informatics*, 50(6), pp. 6.1–6.11. doi: 10.4230/OASICS.SCOR.2016.6.
- Tuite, A. R. *et al.* (2017) 'Stochastic agent-based modeling of tuberculosis in Canadian Indigenous communities', *BMC Public Health*, 17(1), pp. 1–12. doi: 10.1186/s12889-016-3996-7.
- Vila Guilera, J. (2017) 'Analysis and individual-based modelling of the tuberculosis epidemiology in Barcelona. The role of age, gender and origin.', (May). Available at: <https://upcommons.upc.edu/handle/2117/107446>.
- WHO (2020) *Global Tuberculosis Report*, *BMC Public Health*. Available at: <https://ejournal.poltektegal.ac.id/index.php/siklus/article/view/298%0Ahttp://repositorio.unan.edu.ni/2986/1/5624.pdf%0Ahttp://dx.doi.org/10.1016/j.jana.2015.10.005%0Ahttp://www.biomedcentral.com/1471-2458/12/58%0Ahttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&P>.
- Yates, T. A. and Atkinson, S. H. (2017) 'Ironing out sex differences in tuberculosis prevalence', *International Journal of Tuberculosis and Lung Disease*, 21(5), pp. 483–484. doi: 10.5588/ijtld.17.0194.



# ROBUST BIPLLOT ANALYSIS OF NATURAL DISASTERS IN INDONESIA FROM 2019 TO 2021

Hilda Venelia<sup>1</sup>, Khoirin Nisa<sup>1\*</sup>, Rizki Agung Wibowo<sup>1</sup>, Mona Arif Muda<sup>2</sup>

<sup>1</sup> Department of Mathematics, Faculty of Mathematics and Natural Sciences,  
University of Lampung, Lampung, Indonesia

<sup>2</sup>Department of Informatics Engineering, Faculty of Engineering, University of Lampung, Indonesia  
e-mail: khoirin.nisa@fmipa.unila.ac.id

## Abstract

Indonesia is one of the most natural disaster-prone countries in the world, frequently exposed to a range of hazards. Currently, Indonesia has 34 provinces and natural disasters that occur in each province are different, therefore it is necessary to analyze the mapping of natural disasters that often occur in each province to provide scientific analysis for risk management of the natural disasters. One of the quick steps in describing data that can be used is biplot analysis, as biplot analysis can describe a lot of data then summarized it into the form of a two-dimensional graph. The aim of this research is to map 34 provinces in Indonesia based on the incidence of natural disasters from 2019 to 2021 using robust biplot analysis to see which provinces that have a high risk of natural disaster. Based on the result, robust biplot analysis can explain 87,9% of the information on natural disasters in every province in Indonesia. Lampung, Bengkulu, Bangka Belitung, Special Region of Yogyakarta, North Sulawesi, West Sulawesi, Southeast Sulawesi, Gorontalo, East Nusa Tenggara, Bali, Maluku, West Maluku, Papua, and West Papua are provinces that have similar natural disaster characteristics. Flood, tornado and forest and land fires are natural disasters that often occur in Indonesia. The provinces that have the highest risk of flood, landslide, and tornado were West Java, Central Java, and East Java. Then, the provinces with the highest risk of forest and land fires were Aceh and South Kalimantan.

**Kata kunci:** Natural disasters, risk management, biplot.

## INTRODUCTION

Based on Law of Republic of Indonesia number 24 of 2007 concerning Disaster Management, disaster is an event or series of events that threaten and disrupt people's lives and livelihoods, caused by natural and/or non-natural and human factors that result in human casualties, environmental damage, property losses, and psychological impacts. Natural disaster is a catastrophic event with atmospheric, geological, and hydrological origin (e.g., droughts, earthquakes, floods, hurricanes, landslides) that can cause fatalities, property damage and social environmental disruption (Xu, *et al.*, 2016). Every country in the world has been hit by natural disaster, either high, medium, or low risk. Indonesia is one of the countries with a high disaster-prone level in most of its territory. Geographically, Indonesia is an archipelagic country located at the confluence of four tectonic plates, namely Asian Continent plate, Australian Continent plate, Indian Ocean plate and Pacific Ocean. These conditions are potential disaster-prone such as volcanic eruptions, earthquakes, tsunamis, floods and landslides (Arnold, 1986).

Indonesian National Agency for Disaster Management (*Indonesian*: Badan Nasional Penanggulangan Bencana (BNPB)) noted that the annual trend of natural disasters in Indonesia tends to increase. In period of 2010 until 2020, the highest number of annual disasters occurred in 2019, which was 3.814 incidents. Disasters that hit Indonesia are generally caused by hydrometeorology. Flood, landslide, and tornado have dominated the natural disasters that have occurred over the past decade. The number of missing and dead victims annually reaches hundreds to thousands of people. The peak occurred in 2018 with 3.397 disasters which caused the number of victims to reach 6.240 people. That year, Indonesia experienced a series of major disasters, such as Lombok earthquake (West Nusa Tenggara), Palu earthquake and tsunami (Central Sulawesi), and Sunda

Strait tsunami. Meanwhile, from 31 December 2019 to 1 January 2020 there was rain in the Greater Jakarta (i.e. Jakarta Bogor Tangerang Bekasi, abbreviated to Jabodetabek) which resulted in flooding on 1-3 January 2020. The incident killed at least 48 people, and more than 31.000 people were evacuated. Then, floods and landslides also occurred in East Nusa Tenggara in early April 2021 due to Cyclone Seroja which killed 183 people.

Therefore, many provinces in Indonesia are affected by natural disasters in terms of social, economic, environmental, and other aspects. This is due to the lack of readiness of government and society in disaster management before the disaster occurs or better known as Disaster Risk Reduction (DRR). The main objective of DRR is to reduce losses due to disaster impacts by increasing society capacity and reducing exposure and increasing community resilience through preparedness, emergency response, and recovery. The first thing that must be considered in DRR is to map the province in Indonesia based on the level of risk from the natural disaster and find out what natural disasters often occur in each province in Indonesia. This is because, by mapping the provinces in Indonesia based on their level of natural disaster risk, it can be seen which provinces that have high, medium, and low risk of natural disaster. Thus, natural disaster management can be prioritized to provinces that have a high risk of being affected by natural disasters.

There is one of descriptive statistical methods that can be used for mapping purposes, namely biplot analysis. Biplot analysis is one of descriptive statistical methods that can simultaneously present plots of  $n$  observations and  $p$  variables in two-dimensional graph (Jolliffe, 2002). Biplot analysis has been used in various research fields, one can see e.g. (Makhya, *et al.*, 2020; Fitry, *et al.*, 2021; Suryowati, JP and Nasution, 2021; Ghazvini, 2021). However, the classical biplot is sensitive to the presence of outliers in the data, therefore

when data contain outliers it is recommended to use robust biplot.

## METHODOLOGY

The data used in this study is Natural Disasters in Indonesia data from January 1st 2019 until August 14th 2021 obtained from Indonesian National Disaster Mitigation Agency (BNPB). The data consists of  $n = 34$  provinces and  $p = 6$  natural disasters in Indonesia. The variables used in this study are the types of natural disasters, namely Flood (F), Landslide (LS), Earthquake (EQ), Tidal Wave/Abrasion (TW), Forest and Land Fires (FF), and Tornado (T).

The computation of biplot analysis based on Singular Value Decomposition (SVD) (Gabriel, 1971) that can be written as follows:

$$\mathbf{X} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T, \quad (1)$$

where  $\mathbf{X}$  is data matrix with rank  $r$ ,  $\mathbf{U}_{(n \times r)}$  and  $\mathbf{V}_{(p \times r)}$  are columnar orthonormal matrices ( $\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I}_r$ ) and  $\mathbf{\Lambda}_{(r \times r)} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_r})$  with  $\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \dots \geq \sqrt{\lambda_r}$  and  $r \leq p \leq n$ . The element  $\lambda_i, i = 1, 2, \dots, r$  are eigenvalues of  $\mathbf{X}^T \mathbf{X}$  (Johnson and Wichern, 2014). Matrix  $\mathbf{V}$  is a matrix whose columns consist of eigenvectors  $\mathbf{v}_i$  corresponding to  $\lambda_i$  of  $\mathbf{X}^T \mathbf{X}$ . The columns of  $\mathbf{U}$  can be calculated as:

$$\mathbf{U} = \frac{1}{\sqrt{\lambda_i}} \times \mathbf{v}_i.$$

Define  $\mathbf{\Lambda}^\alpha = \text{diag}(\sqrt{\lambda_1^\alpha}, \sqrt{\lambda_2^\alpha}, \dots, \sqrt{\lambda_r^\alpha})$  with  $\alpha \in [0, 1]$  and suppose  $\mathbf{G} = \mathbf{U} \mathbf{\Lambda}^\alpha$ ,  $\mathbf{H} = \mathbf{V} \mathbf{\Lambda}^{1-\alpha}$  then (1) can be written as (Gabriel, 1971):

$$\begin{aligned} \mathbf{X} &= \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T \\ &= (\mathbf{U} \mathbf{\Lambda}^\alpha) (\mathbf{\Lambda}^{1-\alpha} \mathbf{V}^T) \\ &= \mathbf{G} \mathbf{H}^T. \end{aligned}$$

Therefore, element  $(i, j)$  of data matrix  $\mathbf{X}_{(n \times p)}$  can be expressed as:

$$x_{ij} = \mathbf{g}_i^T \mathbf{h}_j$$

where  $\mathbf{g}_i^T, i = 1, 2, \dots, n$  and  $\mathbf{h}_j, j = 1, 2, \dots, p$  are row vectors of  $\mathbf{G}$  and  $\mathbf{H}$ , respectively, with  $r$  elements. Here,  $n$  rows of  $\mathbf{G}$  correspond to rows of  $\mathbf{X}$  and  $p$  rows of  $\mathbf{H}$  correspond to columns of  $\mathbf{X}$ .

Outliers are data points that deviate far from the majority of the data (Filzmoser, 2004). Multivariate outliers can be identified using Mahalanobis distance (Majewska, 2015) as follows:

$$D_i(\mathbf{x}_i, \bar{\mathbf{X}}) = \sqrt{(\mathbf{x}_i - \bar{\mathbf{X}})^T \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{X}})},$$

where  $\mathbf{x}_i$  represents the  $i$ -th object,  $\bar{\mathbf{X}}$  is the mean vector and  $\mathbf{S}$  is the sample covariance matrix (Ghorbani, 2019). If  $D_i^2(\mathbf{x}_i, \bar{\mathbf{X}}) > \chi_{p;1-\alpha}^2$  with  $p$  the number of variable in the data and  $\alpha$  the significant level, then the object can be identified as outlier.

However, not all outliers can be removed from data because it can cause the loss of information contained in data. Furthermore, the covariance matrix also very sensitive to outliers (Islami and Sihombing, 2021; Larasati, *et al.*, 2021). To overcome this problem, biplot analysis can be generated using a robust covariance matrix by estimating eigenvalues and eigenvectors of  $\mathbf{U}$  and  $\mathbf{V}$  such that the predicted results are resistant to outliers (Hawkins, *et al.*, 2001). In this case the calculation of SVD uses robust estimator of mean vector  $\bar{\mathbf{X}}$  and covariance matrix  $\mathbf{S}$ . Here we used the Minimum Covariance Determinant (MCD) estimator which is known as a very robust method for mean vector and covariance matrix estimation (Hubert and Debruyne, 2010). The MCD estimator with the fast-MCD algorithm proposed by Rousseeuw and Van Driessen (1999) can be done with the following steps:

1. Take a random subset of matrix  $\mathbf{X}$ , suppose the subset as  $\mathbf{K}_1$  with the number of elements as much as  $h$ ,

$$h = \frac{(n + p + 1)}{2}$$

where  $n$  and  $p$  are the sample size and the number of variables in the data respectively.

2. Calculate the mean vector  $\bar{\mathbf{X}}_1$  and the covariance matrix  $\mathbf{S}_1$  of  $\mathbf{K}_1$  using:

$$\begin{aligned} \bar{\mathbf{X}}_1 &= \frac{1}{h} \sum_{i \in H} \mathbf{x}_i \\ \mathbf{S}_1 &= \frac{1}{h-1} \sum_{i=1}^h (\mathbf{x}_i - \bar{\mathbf{X}}_1)^T (\mathbf{x}_i - \bar{\mathbf{X}}_1) \end{aligned}$$

3. Calculate the determinant of  $\mathbf{S}_1$ .
4. Calculate the relative distance of each observation to  $\bar{\mathbf{X}}_1$  and covariance  $\mathbf{S}_1$  using Mahalanobis distance.
5. Sort the observation by the distance of Mahalanobis, from the smallest to largest.
6. Take the elements of  $h$  observations with the smallest distance based on step 5 to become subset of  $\mathbf{K}_2$ , repeat steps 2 until 5 so that it is found the subsets are converge and have the smallest determinant of the covariance matrix, namely:

$$|\mathbf{S}_{n+1}| < |\mathbf{S}_n|$$

7. Based on the elements of  $h$ , the next data is weighted as:

$$w_i = \begin{cases} 1, & (\mathbf{x}_{ij} - \bar{\mathbf{X}})^T \mathbf{S}^{-1} (\mathbf{x}_{ij} - \bar{\mathbf{X}}) < \chi_{p,1-\alpha}^2 \\ 0, & \text{others} \end{cases}$$

8. Based on the weighted above, the fast-MCD estimators are:

$$\bar{\mathbf{X}}_{MCD} = \frac{\sum_{i=1}^n w_i \mathbf{x}_{ij}}{\sum_{i=1}^n w_i}$$

$$\mathbf{S}_{MCD} = \frac{\sum_{i=1}^n w_i (\mathbf{x}_{ij} - \bar{\mathbf{X}}_{MCD})^T (\mathbf{x}_{ij} - \bar{\mathbf{X}}_{MCD})}{(\sum_{i=1}^n w_i) - 1}$$

Robust biplot analysis in this research was carried out using R software, the analysis procedure can be described as follows:

1. data screening to detect outliers by using Mahalanobis distance;
2. performing robust biplot analysis based on the covariance matrix  $\mathbf{S}_{MCD}$  using *robustbase* package;
3. plotting data using *ggbiplot* package;

4. interpretation of the result.

## RESULT & DISCUSSION

In this research, the first step was to detect outliers in data based on the robust Mahalanobis distance. The plot of the robust squared Mahalanobis distances and the chi-square quantiles is shown in Figure 1.

The result showed that there were ten provinces that were identified as outliers as can be seen in Figure 1. The list of data that identified as outliers is presented in the following table 1.

In Table 1, Aceh, North Sumatra, West Sumatra, South Kalimantan, and South Sulawesi are provinces that experience flood and tornado with high risk, especially in West Java, Central Java, and East Java that experience flood, landslide, and tornado with very high risk than other provinces. While East Kalimantan and Bali are provinces that experience flood and tornado with low risk than others.

Based on the report above it can be concluded that the data of natural disasters in Indonesia from 2019 until 2021 contains outliers. Therefore, it is appropriate to use robust biplot analysis in this research. The resulted robust biplot describes the characteristics of object (Provinces) against the variables (Natural Disasters) as can be seen in the following figure.

The important thing that needs to be considered in biplot analysis is how good the biplot can explain the information

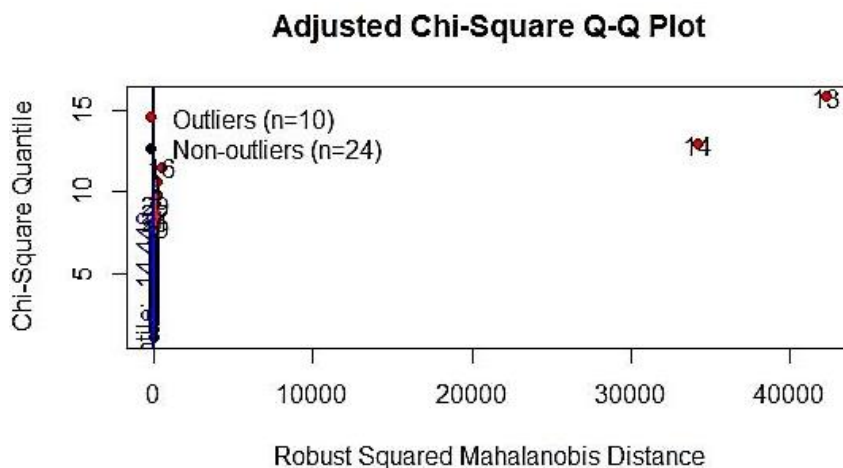


Figure 1. Outlier detection based on robust Mahalanobis distance

Table 1. Provinces that are considered as outliers and their number of natural disasters

Province	Flood	Landslide	Earthquake	Tidal Wave	Forest and Land Fires	Tornado
Aceh	217	33	4	13	245	193
North Sumatra	108	22	1	1	27	96
West Sumatra	133	68	1	7	24	164
West Java	452	883	26	3	83	895
Central Java	446	788	3	11	95	791
East Java	344	99	30	2	195	394
South Kalimantan	110	17	0	3	168	112
East Kalimantan	41	37	0	1	105	13
Bali	14	48	5	4	14	79
South Sulawesi	119	54	1	7	28	159

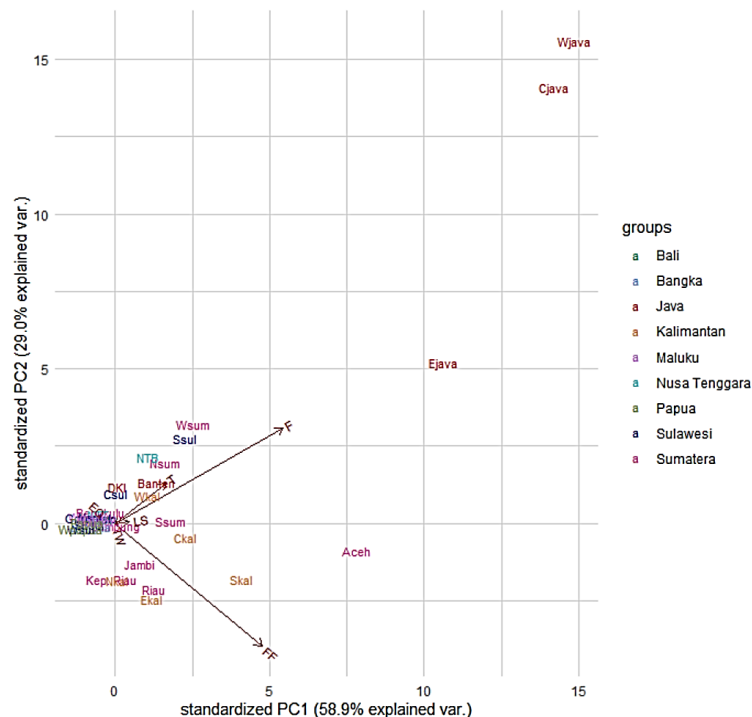


Figure 2. Robust biplot charts of natural disasters.

contained in data. Based on Figure 2, it can be explained that the information in component 1 was 58,9% and component 2 was 29%. Therefore, the cumulative information that can be explained by biplot against data based on two components was 87,9%. It means that the resulted biplot was very capable in describing the characteristics of natural disaster data in Indonesia. Furthermore, there are four information obtained from biplot in Figure 2 as described as follows.

### Proximity Between Objects

This information can be used as a guide to identifying objects that have

similar characteristics to other objects. Two objects that have similar characteristics will be described as two points with adjacent positions. Based on Figure 2, West Sumatra and South Sulawesi are two provinces with similar characteristics. West Nusa Tenggara and North Sumatra are two provinces with similar characteristics. DKI Jakarta and Central Sulawesi are two provinces with similar characteristics. Then, Banten and West Kalimantan are two provinces with similar characteristics. Also with Riau and East Kalimantan have similar characteristics. Jambi, Riau Islands, and North Kalimantan are three provinces with similar characteristics. Meanwhile,

Lampung, Bengkulu, Bangka Belitung, Special Region of Yogyakarta, North Sulawesi, West Sulawesi, Southeast Sulawesi, Gorontalo, East Nusa Tenggara, Bali, Maluku, West Maluku, Papua, and West Papua form their own group and have similar characteristics.

### Variability of Variables

This information is used to see if there are variables that have almost the same diversity. Variables with small variance are represented by short vectors, whereas if the variance is high, they are represented by long vectors. Based on Figure 2, flood and forest fire are variables with the highest variance compared to other variables because they have the longest vector. It means that the level of risk of flood and forest fires varies widely in each province in Indonesia. Then, tidal wave is variable with the smallest variance represented by the shortest vector. This means, tidal wave was a natural disaster with the lowest risk in every province in Indonesia from 2019 until August 2021.

### Correlation Between Variables

Correlation between variables shows how one variable is related to another. In biplot, variables are presented as vector lines. If two vectors form an acute angle ( $< 90^\circ$ ) or have the same direction, then it can be stated that the two variables are positively correlated. Whereas if two vectors form obtuse angles ( $> 90^\circ$ ) or have opposite directions, it can be stated that the two variables are negatively correlated. Then, if the angle formed between two vectors is a right angle, then it can be stated that the two variables are not correlated.

Based on Figure 2, flood has positive correlation with forest and land fires, tornado, and landslide. This means that from 2019 until August 2021, the level of flood risk in each province in Indonesia was directly proportional to the level of risk of forest and land fires, tornado, and landslide. Earthquake has negative correlation with flood, landslide, and forest and land fires which means from 2019 until August 2021, the level of earthquake risk was inversely

proportional to the level of risk of flood, landslide, and forest and land fires. But earthquake, tornado, and forest and land fires have no correlation.

### Relative Position of Object to Variable

This information is used to see the advantages of each object. Objects that are in the same direction as the variable vector, indicate that the value of the object is above the average, if it is in the opposite direction, it means that the value is below the average, if it is almost in the middle, it means that the value is close to the average. Based on Figure 2, information about natural disasters in Indonesia from 2019 until August 2021 shows that West Java, Central Java, and East Java are provinces with the highest risk of flood, tornado, and landslide than other provinces. Whereas, West Sumatra, North Sumatra, South Sumatra, Banten, DKI Jakarta, West Kalimantan, West Nusa Tenggara, South Sulawesi, and Central Sulawesi are provinces with a medium risk of flood and tornado.

Aceh and South Kalimantan were at the highest risk of forest and land fires. Meanwhile, South Sumatra, Jambi, Riau, Riau Islands, Central Kalimantan, North Kalimantan, and East Kalimantan have a medium risk of forest and land fires and a lower risk of earthquake. Meanwhile, Lampung, Bengkulu, Bangka Belitung, Special Region of Yogyakarta, North Sulawesi, West Sulawesi, Southeast Sulawesi, Gorontalo, East Nusa Tenggara, Bali, Maluku, West Maluku, Papua, and West Papua are provinces with a medium risk of earthquake and a lower risk of forest and land fires.

### CONCLUSION

Based on results and discussion, there are several points that can be concluded as follow:

1. The information obtained from robust biplot analysis is 87,9% which means very good to describe the characteristics of natural disasters in Indonesia.
2. Flood, tornado, and forest and land fires are natural disasters that often occur in Indonesia. Meanwhile, tidal wave is a

natural disaster that is very rare occurrence in Indonesia.

3. Lampung, Bengkulu, Bangka Belitung, Special Region of Yogyakarta, North Sulawesi, West Sulawesi, Southeast Sulawesi, Gorontalo, East Nusa Tenggara, Bali, Maluku, West Maluku, Papua, and West Papua are provinces that have similar natural disaster characteristics and have medium risk of earthquake.
4. West Java, Central Java, and East Java have the highest risk of being affected by flood, tornado, and landslide. Meanwhile, Aceh and South Kalimantan are provinces at high risk of being affected by forest and land fires.

Based on these information, it is recommended for further research to examine the factors and impacts of natural disasters that occur in Indonesia.

## REFERENCES

- Arnold, E.P. 1986. Series on Seismology Volume V: Indonesia. South-East Asia Association of Seismology and Earthquake Engineering.
- Filzmoser, P. 2004. *Multivariate Outlier Detection Method*. Austria: Department of Statistics and Probability Theory.
- Filzmoser, P. and Gregorich, M. 2020. Multivariate Outlier Detection in Applied Data Analysis: Global, Local, Compositional, and Cellwise Outliers. *Mathematical Geosciences*, Vol. 52 (8), 1049-1066.
- Fitry, R., Rostini, N., Hersanti and Zubair, A. 2021. Biplot and Correlation Analysis of 189 Superior F2 Genotypes Chili-Unpad in Indonesia. *Advances in Biological Sciences Research*, Vol. 9, 10-14.
- Gabriel, K. R. 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, Vol. 58 (3), 453-467.
- Ghazvini, H., Bagherikia, S., Pour-Aboughadareh, A., Sharifalhossaini, M., Razavi, S.A., Mohammadi, S., Ghasemi, K. M., Fathihafshejani, A. and Khakizade, G. 2021. GGE Biplot Analysis of Promising Barley Lines in the Cold Regions of Iran. *Journal of Crop Improvement*, Vol. 35, 1-12.
- Ghorbani, H. 2019. Mahalanobis Distance and Its Application for Detecting Multivariate Outliers. *Facta Universitatis: Mathematics and Informatics*, Vol. 34, 583.
- Hawkins, D. M., Liu, L. and Young, S. S. 2001. Robust Singular Value Decomposition. *Linear Algebra and its Applications*, Vol. 417 (122), 370-380.
- Hubert, M. and Debruyne, M. 2010. Minimum covariance determinant. *Wiley Reviews: Computational Statistics*, Vol. 2 (1), 36-43.
- Islami, R. L. and Sihombing, P. R. 2021. Application Biplot and K-Medians Clustering to Group Export Destination Countries of Indonesia's Product. *Advance Sustainable Science, Engineering and Technology*, Vol. 3 (1), 0210105.
- Johnson, R. and Wichern D. 2014. *Applied Multivariate Statistical Analysis 6th ed.* England: Pearson Education Limited.
- Jolliffe, I. T. 2002. *Principal Component Analysis, Second Edition*. New York: Springer-Verlag.
- Larasati, S.D.A., Nisa, K. and Herawati, N. 2021. Robust Principal Component Trimmed Clustering of Indonesia Provinces Based on Human Development Index Indicators. *Journal of Physics: Conference Series*, Vol. 1751 (1).
- Mahalanobis, P. C. 1936. On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, Vol. 2 (1), 49-55.
- Majewska, J. 2015. Identification of Multivariate Outliers Problems and Challenges of Visualization Methods. *Informatyka i Ekonometria* 247, 69-83.
- Mattjik, A. A. and Sumertajaya, I. M. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. Bogor: IPB Press.

- Rousseeuw, P. J. and Driessen, K. V. 1999. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, Vol. 41 (3), 212-223.
- Suryowati, K, Titah, J.P., and Nasution, N. 2021. Application Biplot Analysis on Mapping of Non-Convertible Diseases in Indonesia. *Parameter: Journal of Statistics*, Vol. 1 (2), 11-20.
- Syarief, M., Nisa, K., Larasati, S., Suropto and Suprihatin, A. 2020. Government policy in the case of provision of commercial bank loans based on business fields per Sumatra region. *TEST Engineering and Management*, Vol. 82, 6843-6851.
- Xu, J., Wang, Z., Shen, F., Ouyang, C. and Tu, Y. 2016. Natural disasters and social conflict: A systematic literature review. *Int. J. Disaster Risk Reduct*, Vol. 17, 38-48.



# Petunjuk Penulisan

## JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

Naskah dikirim dalam bentuk *softcopy* ke alamat email [pppm@stis.ac.id](mailto:pppm@stis.ac.id) disertai dengan daftar riwayat hidup ringkas penulis. Format naskah mengacu pada Petunjuk Penulisan Naskah berikut:

Naskah dibuat menggunakan *Microsoft Office Word* 2010. Seluruh bagian dalam naskah diketik dengan huruf *Times New Roman*, ukuran 12, spasi 1,5, ukuran kertas A4 dan margin 2 cm untuk semua sisi, serta jumlah halaman 15-20. Untuk kepentingan penyuntingan naskah, seluruh bagian naskah (termasuk tabel, gambar dan persamaan matematika) dibuat dalam format yang dapat disunting oleh editor.

Gaya penulisan naskah untuk Jurnal Aplikasi Statistika dan Komputasi Statistik ditulis dalam Bahasa Indonesia dengan gaya naratif. Pembabakan dibuat sederhana dan sedapat mungkin menghindari pembabakan bertingkat. Tabel dan gambar harus mencantumkan sumber jika dari data sekunder. Tabel, gambar dan persamaan matematika diberi nomor secara berurut sesuai dengan kemunculannya. Semua kutipan dan referensi dalam naskah harus tercantum dalam daftar pustaka, dan sebaliknya sumber bacaan yang tercantum dalam daftar pustaka harus ada dalam naskah. Format sumber: Nama Penulis dan Tahun. Nomor dan judul tabel diletakkan di bagian atas tabel dan dicetak tebal, sedangkan nomor dan judul gambar diletakkan di bagian bawah gambar dan dicetak tebal.

### Bagian naskah berisi:

**Judul.** Judul tidak melebihi 12 kata dalam Bahasa Indonesia.

**Data Penulis.** Berisi nama lengkap semua penulis tanpa gelar, asal institusi, dan alamat email.

**Abstrak.** Ditulis dalam Bahasa Inggris dan Bahasa Indonesia, maksimum 100 kata untuk masing-masing abstrak dan berisikan tiga hal yaitu topik yang dibahas, metodologi yang dipergunakan dan hasil yang didapatkan.

**Kata Kunci.** Berisi kata atau frasa (maksimum 5 subjek) yang sering dipergunakan dalam naskah dan dianggap mewakili dan atau terkait dengan topik yang dibahas.

**Pendahuluan.** Memuat latar belakang, studi sebelumnya yang relevan, permasalahan ataupun hipotesis yang akan diuji dalam penelitian, ruang lingkup penelitian, serta tujuan dari penelitian.

**Metodologi** terdiri atas:

- a. **Tinjauan Referensi.** Bagian ini menguraikan landasan konseptual dari tulisan dan berisi alasan teoritis mengapa pertanyaan penelitian dalam artikel diajukan. Di samping itu penulis dapat mengutip studi yang relevan sebelumnya untuk melengkapi justifikasi mengenai kerangka pikir penelitian.
- b. **Metode Analisis.** Bagian ini berisi informasi teoritis dan teknis yang cukup memadai untuk pembaca dapat mereproduksi penelitian dengan baik termasuk di dalamnya uraian mengenai jenis dan sumber data serta variabel yang digunakan. Dalam hal keperluan verifikasi hasil, editor dan mitra bestari (*reviewer*) berhak meminta data mentah (*raw data*) yang digunakan penulis.

**Hasil dan Pembahasan.** Tuliskan hasil yang didapat berdasarkan metode yang digunakan disertai analisis terhadap variabel-variabelnya . Dapat disajikan berupa tabel, gambar, hasil pengujian hipotesis dengan disertai uraian analitis yang mengangkat poin-poin penting berdasarkan konsepsi teoritisnya.

**Kesimpulan dan Saran.** Bagian ini memuat kesimpulan dari hasil dan implikasinya secara akademis, dan saran yang dapat diberikan berdasarkan temuan dari pembahasan. Bagian ini juga memuat keterbatasan penelitian dan kemungkinan penelitian lanjutan yang dapat dilakukan dengan penggunaan/pengembangan variabel, metode analisis ataupun cakupan wilayah penelitian lainnya.

**Daftar Pustaka.** Daftar pustaka disusun berdasarkan urutan abjad dengan ketentuan sebagai berikut:

**Publikasi Buku**

1. Penulis satu orang  
Enders, Walter. 2010. *Applied Econometric Time Series, Third Edition*. New Jersey: Wiley.
2. Penulis dua orang  
Pyndick, Robert. S. dan Rubinfeld, Daniel L. 2009. *Microeconomics, Seventh Edition*. New Jersey: Pearson Education.
3. Penulis tiga orang  
Fotheringham, A. S., Brunson, C, dan Charlton, M. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex: John Wiley & Sons.

**Artikel dalam jurnal**

Romer, P. 1993. Idea Gaps and Object Gaps in Economic Development. *Journal of Monetary Economics*, Vol. 32 (3), 543–573.

**Artikel online**

Woodward, Douglas P. 1992. Locational Determinants of Japanese Manufacturing Start-Ups in the United States. *Southern Economic Journal*, Vol. 58 (3), 690-708.  
<http://www.jstor.org/discover/10.2307/1059836> (Diakses 1 September, 2014).

**Buku yang ditulis oleh lembaga atau organisasi**

BPS. 2009. *Analisis dan Penghitungan Tingkat Kemiskinan 2008*. Jakarta: BPS.

**Kertas kerja (working papers)**

Edwards, S. 1990. Capital Flows, Foreign Direct Investment, and Debt-Equity Swaps in Developing Countries. *NBER Working Paper*, 3497.

**Makalah yang direpresentasikan**

Zhang, Kevin H. 2006. Foreign Direct Investment and Economic Growth in China: A Panel Data Study for 1992-2004. *Conference of WTO, China, and Asian Economies*. Beijing.

**Karya yang tidak dipublikasikan**

Hartono, Djoni. 2002. Analisis Dampak Kebijakan Harga Energi terhadap Perekonomian dan Distribusi Pendapatan di DKI Jakarta: Aplikasi Model Komputasi Keseimbangan Umum (Computable General Equilibrium Model). *Tesis*. Jakarta.

**Artikel di koran, majalah, dan periodik sejenis**

Reuters. (2014, September 17). Where is Inflation?. *Newsweek*.