JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 13, NOMOR 1, JUNI 2021 ISSN 2086 – 4132
AKREDITASI NOMOR: 747/Akred/P2MI-LIPI/04/2016

Kajian Usaha Rumah Tangga Budidaya Bandeng di Sulawesi Selatan Tahun 2013
MUHAMAD SAIFUL HADI dan PUTRI LYDIA ELTHEOFANY S.

Pemodelan Sebaran *Total Dissolved Solid* Menggunakan Metode *Mixed Geographically Weighted Regression*
DADAN KUSNANDAR, KUSNANDAR, NAOMI NESSYANA DEBATARAJA, dan SITI FITRIANI

Penerapan Peta Kendali Atribut Klasik dan Peta Kendali NP Bayes pada Produk Cacat Air Minum Asri di CV. Multi Rejeki Selaras Payakumbuh
FERRA YANUAR, MUTIARA FARA NABILLA, dan IZZATI RAHMI

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

Pemodelan Kerugian Bencana Banjir Akibat Curah Hujan Ekstrem Menggunakan EVT dan *Copula*
IAN SURYA PRAYOGA dan ATINA AHDIKA

Pengelompokkan Provinsi di Indonesia Berdasarkan Jumlah Kasus Covid-19 dan Fasilitas Kesehatan
MEINISA FADILLAH RAHMI, PAULUS SATRIA PRASETYO, RATIH NURHABIBAH, RIZKY PERDANA, dan WA ODE ZUHAYENI MADJIDA

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.

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 kehadirat 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 NOMOR: 747/Akred/P2MI-LIPI/04/2016

DAFTAR ISI

Pengantar Redaksi.....iii

Daftar Isi..... iv

Abstrak..... v-xii

Kajian Usaha Rumah Tangga Budidaya Bandeng di Sulawesi Selatan Tahun 2013
Muhamad Saiful Hadi dan Putri Lydia Eltheofany S.....1-8

Pemodelan Sebaran *Total Dissolved Solid* Menggunakan Metode *Mixed Geographically Weighted Regression*
Dadan Kusnandar, Kusnandar, Naomi Nessyana Debataraja, dan Siti Fitriani.....9-16

Penerapan Peta Kendali Atribut Klasik dan Peta Kendali NP Bayes pada Produk Cacat Air Minum Asri di CV. Multi Rejeki Selaras Payakumbuh
Ferra Yanuar, Mutiara Fara Nabilla, dan Izzati Rahmi 17-24

***Small Area Estimation with Excess Zero* (Studi Kasus: Angka Kematian Bayi di Pulau Jawa)**
Nofita Istiana.....25-34

Pemodelan Kerugian Bencana Banjir Akibat Curah Hujan Ekstrem Menggunakan EVT dan *Copula*
Ian Surya Prayoga dan Atina Ahdika 35-46

Pengelompokkan Provinsi di Indonesia Berdasarkan Jumlah Kasus Covid-19 dan Fasilitas Kesehatan
Meinisa Fadillah Rahmi, Paulus Satria Prasetyo, Ratih Nurhabibah, Rizky Perdana, dan Wa Ode Zuhayeni Madjida 47-56

JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

(*Journal of Statistical Application & Statistical Computing*)

ISSN 2086 – 4132 **Volume 13, Nomor 1, Juni 2021**

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

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%.

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

DDC: 315.98

Dadan Kusnandar, Kusnandar, Naomi Nessyana Debataraja, 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 Bisquare 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 rataan 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

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

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

JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

(*Journal of Statistical Application & Statistical Computing*)

ISSN 2086 – 4132 **Volume 13, Nomor 1, Juni 2021**

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 Nessyana Debataraja, 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

DDC: 315.98

Keywords: Extreme rainfall, floods, Extreme Value Theory, Copula

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

KAJIAN USAHA RUMAH TANGGA BUDIDAYA BANDENG DI SULAWESI SELATAN TAHUN 2013

Muhammad Saiful Hadi[1] dan Putri Lydia Eltheofany S[2]

[1]Badan Pusat Statistik Kabupaten Yakuhimo, [2]Badan Pusat Statistik Kabupaten Mimika
e-mail: [1]11.6802@stis.ac.id, [1]putri.lydia@bps.go.id

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%.

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

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*

PENDAHULUAN

Indonesia saat ini menjadi negara produsen perikanan terbesar kedua di dunia setelah China, sejak tahun 2009-2013 Indonesia menempati posisi kedua (Direktorat Jenderal Perikanan Budidaya KKP, 2015). Persentase kenaikan rata-rata produksi Indonesia merupakan kenaikan tertinggi dibandingkan dengan 10 besar negara penghasil perikanan budidaya dunia. Persantese kenaikan rata-rata produksi Indonesia sebesar 27,84%. Sedangkan China yang merupakan negara terbesar hanya sebesar 5,29% kenaikan rata-rata produksinya. (Direktorat Jenderal Perikanan Budidaya KKP, 2015).

Pada tahun 2013 pertumbuhan Produk Domestik Bruto (ADHK tahun 2000) subsektor perikanan mencapai 6,86%. Pertumbuhan tersebut melebihi besarnya pertumbuhan PDB Indonesia pada tahun 2013, yaitu sebesar 5,73%, dan jika tanpa minyak dan gas 6,20% (Badan Pusat Statistik, 2014). Indonesia memiliki beberapa komoditas yang menjadi andalan dalam subsektor perikanan budidaya yang dikembangkan dan menjadi fokus dalam peningkatan produksi perikanan budidaya. Perikanan budidaya tersebut diantaranya udang, rumput laut, bandeng, kerapu, kakap, nila, mas, lele, patin, dan gurame. Secara total produksi, Indonesia berada di posisi kedua sebagai produsen ikan dari hasil budidaya. Salah satu komoditas unggulan Indonesia yang berpotensi dapat bersaing dengan negara-negara lain yaitu bandeng.

Produksi bandeng dunia pada tahun 2013 mencapai satu juta ton lebih. Setiap tahunnya rata-rata produksi bandeng dunia meningkat sebesar 6,99%. Posisi Indonesia pada tahun 2013 adalah produsen nomer satu dengan 575,256 ton atau berkontribusi lebih dari 50% terhadap produksi bandeng dunia. Kenaikan rata-rata bandeng Indonesia setiap tahunnya sebesar 10,90% (Direktorat Jenderal Perikanan Budidaya KKP, 2014).

Provinsi Sulawesi Selatan selama 2006-2012 berturut-turut menempati posisi pertama, dari tahun 2006-2012 memiliki rata-rata kontribusi produksi bandeng sekitar 20% terhadap total seluruh produksi ikan bandeng di Indonesia. Menurut data Sensus Pertanian (ST) tahun 2013, rumah tangga yang melakukan usaha budidaya ikan menurut jenis budidaya tambak/air payau tertinggi dari Sulawesi Selatan yaitu sebanyak 31.695 rumah tangga. Dari jumlah tersebut, rumah tangga yang membudidayakan bandeng sebanyak sekitar 23.720 rumah tangga. Sekitar 75% masyarakat di Sulawesi Selatan melakukan usaha budidaya bandeng.

Tabel 1. Hasil Produksi Bandeng 2010

Provinsi	Pangsa (%)	Produksi (ton)
Sulawesi Selatan	18,54	78.181
Jawa Timur	18,24	76.937
Jawa Barat	15,68	66.146
Jawa Tengah	13,56	57.201
Sulawesi Tenggara	7,78	32.812
Nanggroe Aceh Darussalam	4,85	20.455
Kalimantan Timur	4,11	17.317
Sulawesi Barat	3,36	14.159
Banten	2,62	11.071
Kalimantan Selatan	2,43	10.239

Sumber: Sentra Produksi Perikanan Budidaya 2010

Berdasarkan uraian diatas, provinsi Sulawesi Selatan dijadikan objek penelitian. Bertujuan untuk mengetahui karakteristik, variabel-variabel yang memengaruhi produksi, dan kondisi usaha pada provinsi Sulawesi Selatan. Penelitian ini menggunakan pendekatan fungsi produksi *Cobb-Douglas* (Situmorang, 2007)

KAJIAN PUSTAKA DAN METODOLOGI

Menurut Joerson (2003), produksi merupakan hasil akhir dari proses atau aktifitas ekonomi dengan memanfaatkan beberapa masukan atau input, dengan pengertian ini dapat dipahami bahwa kegiatan produksi adalah mengkombinasikan berbagai input atau masukan untuk menghasilkan output.

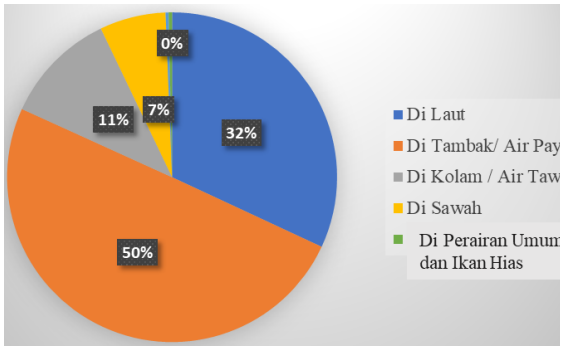
Hubungan teknis antara input dan output tersebut dalam bentuk persamaan, tabel atau grafik merupakan fungsi produksi. Jadi, fungsi produksi adalah suatu persamaan yang menunjukkan jumlah maksimum output yang dihasilkan dengan kombinasi tertentu. (Heryansyah dkk, 2013).

Untuk menganalisis karakteristik dan kondisi rumah tangga usaha budidaya bandeng di provinsi Sulawesi Selatan pada tahun 2013, peneliti mengajukan beberapa hipotesis penelitian yakni 1) variabel tenaga kerja, luas lahan, benih/bibit, pupuk dan obat-obatan berpengaruh signifikan positif terhadap jumlah produksi, 2) skala pengembalian mengalami *increasing return to scale* (IRTS), 3) skala ekonomis dari usaha budidaya bandeng lebih dari 50% rumah tangga.

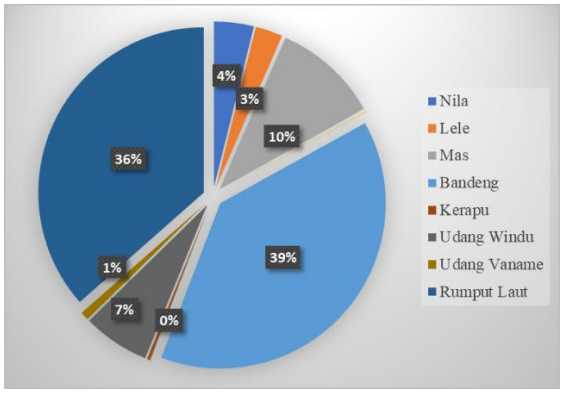
Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu *raw data* dari sensus pertanian 2013. Kuisoner survei rumah tangga usaha budidaya ikan tahun 2014 di provinsi Sulawesi Selatan yang diperoleh dari Subdirektorat Statistik Perikanan, Badan Pusat Statistik Republik Indonesia. Pengumpulan yang dilakukan yaitu survei dari populasi seluruh rumah tangga yang melakukan usaha budidaya ikan bandeng. Sementara itu, metode analisis yang digunakan dalam penelitian ini adalah analisis deskriptif untuk menjelaskan karakteristik rumah tangga di provinsi Sulawesi Selatan pada tahun 2013, serta analisis inferensia regresi linier berganda untuk memeriksa pengaruh tenaga kerja, luas lahan, benih/bibit, pupuk dan obat-obatan terhadap jumlah produksi, untuk menjelaskan kondisi skala pengembalian dan skala ekonomis kondisi rumah tangga yang melakukan usaha budidaya bandeng di Sulawesi Selatan.

HASIL DAN PEMBAHASAN

Dari gambar 1 dapat diketahui bahwa persentase rumah tangga usaha budidaya ikan menurut jenis budidaya ikan di provinsi Sulawesi Selatan pada tahun 2013, data bersumber dari sensus pertanian (ST13) yang diadakan oleh BPS. Jumlah seluruh rumah tangga usaha budidaya ikan pada provinsi Sulawesi Selatan sebesar 62.050 rumah tangga. Dapat diketahui bahwa tertinggi terdapat pada usaha di tambak atau air payau yang persentasenya mencapai 50% dari seluruh jumlah rumah tangga budidaya perikanan, setelah itu terdapat 32% rumah tangga melakukan usaha budidaya perikanan di laut, yang menunjukkan bahwa provinsi Sulawesi Selatan merupakan daerah kelautan, karena lebih dari 80% rumah tangga budidaya perikanan melakukan usaha pada perairan asin ataupun payau.



Gambar 1. Persentase Jumlah Rumah Tangga Usaha Budidaya Ikan Menurut Tempat Budidaya Ikan di Provinsi Sulawesi Selatan pada Tahun 2013Grafik Deret Waktu Kasus Aktif COVID-19



Gambar 1. Persentase Jumlah Rumah Tangga Usaha Budidaya Bukan Ikan Hias Menurut Jenis Ikan Utama di Provinsi Sulawesi Selatan pada Tahun 2013

Dari gambar 2 menunjukkan persentase banyaknya rumah tangga yang melakukan usaha budidaya perikanan menurut jenis ikan utama tanpa ikan hias di provinsi Sulawesi Selatan pada tahun 2013. Didapat persentase tertinggi terdapat pada bandeng dan yang kedua rumput laut,

dengan masing-masing 39% dan 36%. Kedua komoditas unggulan tersebut merupakan komoditas yang diunggulkan di Indonesia. Rumah tangga budidaya perikanan yang melakukan usaha pada komoditas bandeng dan rumput laut, merupakan terbanyak dari provinsi Sulawesi Selatan dibanding semua provinsi di Indonesia. Kedua komoditas ini juga merupakan komoditas yang diunggulkan Indonesia, karena dunia sangat bergantung dan dipengaruhi oleh produksi bandeng dan rumput laut Indonesia.

Setelah itu, terdapat persentase rumah tangga budidaya perikanan di komoditi lain, yaitu ikan mas, udang windu, ikan nila,ikan lele, dan udang vaname masing-masing sebesar 10%, 7%, 4%, 3%, dan 1%. Persentase tersebut sangat berbeda jauh dengan sebelumnya, karena kebanyakan usaha perikanan ini menggunakan air tawar, sedangkan daerah provinsi Sulawesi Selatan didominasi air asin atau air laut dan air payau.

Usaha budidaya bandeng di Sulawesi Selatan pada tahun 2013 memiliki rata-rata produksi sekitar 1191,71 kg, rata-rata luas lahan yang produktif untuk melakukan usaha budidaya sekitar 22.759,13 m2, rata-rata biaya untuk membeli benih/bibit sekitar Rp 724.590,00, rata-rata biaya untuk pupuk dan obat-obatan sekitar Rp 2.017.840,00, dan rata-rata upah tenaga kerja sekitar Rp 2.137.650,00.

Persamaan yang terbentuk sebagai berikut:

$$y_i = -1{,}42 + 0{,}20x*_{1i} + 0{,}24x*_{2i} + 0{,}28x*_{3i} + 0{,}37x*_{4i}$$

Keterangan: *) signifikan pada taraf uji 5%

y_i : logaritma natural dari produksi (kg) rumah tangga ke-i

x_{1i} : logaritma natural dari luas lahan (m^2) rumah tangga ke-i

x_{2i} : logaritma natural dari biaya benih/bibit (000 Rp) rumah tangga ke-i

x_{3i} : logaritma natural dari biaya pupuk & obat-obatan (000 Rp) rumah tangga ke-i

x_{4i} : logaritma natural dari biaya tenaga kerja (000 Rp) rumah tangga ke-i

Hasil estimasi model menunjukkan bahwa *prob* uji simultan F statistik bernilai 0,000 yang berarti seluruh variabel bebas dalam model memengaruhi variabel tak bebas secara signifikan. Nilai *R-squared* sebesar 0,782 menunjukkan bahwa seluruh variabel input mampu menjelaskan keragaman jumlah produksi (output) sebesar 78,2%, sedangkan sisanya 22,8% dijelaskan oleh variabel lain diluar model. Sementara itu, secara signifikan secara parsial kelima variabel input berpengaruh signifikan terhadap output.

Tabel 2. Hasil Estimasi Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.418832	0.146543	-9.682002	0.0000
X1	0.196725	0.021659	9.082766	0.0000
X2	0.240172	0.020509	11.71052	0.0000
X3	0.277310	0.017900	15.49257	0.0000
X4	0.367682	0.021557	17.05608	0.0000

R-squared	0.782040	Mean dependent var	6.396476
Adjusted R-squared	0.781281	S.D. dependent var	1.227079
S.E. of regression	0.573873	Akaike info criterion	1.731505
Sum squared resid	378.4001	Schwarz criterion	1.753390
Log likelihood	-994.0785	Hannan-Quinn criter.	1.739765
F-statistic	1030.653	Durbin-Watson stat	1.420082
Prob(F-statistic)	0.000000		

Selanjutnya intersep menggambarkan indeks efisiensi produksi. Jika saat semua variabel independen bernilai nol (0) atau tidak ada penggunaan input, maka produksi akan menurun sebesar 1,42. Berarti jika suatu usaha tidak dijalankan, maka akan mengalami kerugian secara ekonomis. Kemudian, variabel luas lahan terhadap produksi menunjukkan nilai koefisien sebesar 0,20 yang berarti bahwa setiap peningkatan/pertambahan luas lahan sebesar 1% akan diikuti oleh peningkatan produksi sebesar 0,20% dengan asumsi *ceteris paribus*, variabel ini memberikan kontribusi terendah terhadap produksi bandeng.

Variabel biaya benih/bibit terhadap produksi menunjukkan nilai koefisien sebesar 0,24 yang berarti bahwa setiap peningkatan/pertambahan biaya benih/bibit sebesar 1% akan diikuti oleh peningkatan produksi sebesar 0,24% dengan asumsi *ceteris paribus*. Variabel biaya pupuk &

obat-obatan terhadap produksi menunjukkan nilai koefisien sebesar 0,28 yang berarti bahwa setiap peningkatan/pertambahan biaya pupuk & obat-obatan sebesar 1% akan diikuti oleh peningkatan produksi sebesar 0,28% dengan asumsi *ceteris paribus*. Variabel biaya tenaga kerja terhadap produksi menunjukkan nilai koefisien sebesar 0,38 yang berarti bahwa setiap peningkatan/pertambahan biaya tenaga kerja sebesar 1% akan diikuti oleh peningkatan produksi sebesar 0,38% dengan asumsi *ceteris paribus*, variabel ini memberikan kontribusi tertinggi terhadap produksi bandeng. Pernyataan diatas sejalan dengan penelitian (Kumalasari, 2004) yang menyatakan tenaga kerja, luas lahan, benih, pakan, pupuk, dan pestisida berpengaruh positif terhadap produksi tambak bandeng di Kabupaten Cilacap.

Setelah persamaan terbentuk, dapat diketahui jumlah *slope* (nilai yang menunjukkan besarnya konstribusi) dari tiap variabel. Hasil dari penjumlahan tersebut yaitu sebesar 1,081889 yang nilainya lebih dari 1, menyatakan bahwa kondisi usaha rumah tangga budidaya ikan bandeng di Sulawesi Selatan pada tahun 2014 berada pada IRTS (*Increasing Return to Scale*). Berarti bahwa rumah tangga yang melakukan usaha budidaya ikan bandeng di Sulawesi Selatan dalam kondisi bagus jika dikembangkan usahanya, karena jika semua input dilipatgandakan, output yang dihasilkan lebih dari dua kali lipat. Hasil ini sejalan dengan penelitian (Kumalasari, 2004) yang menyatakan hasil usaha tambak bandeng di Kabupaten Cilacap berada dalam kondisi *Increasing Return to Scale*.

Menurut (Christensen and Green, 1976) perhitungan translog biaya mempunyai bentuk umum sebagai berikut:

$$\ln C_i = \alpha_0 + \alpha_Q \ln Q_i + \frac{1}{2}\gamma_{QQ}(lnQ)^2 + \sum_{j=1}^{p} \alpha_j \, lnw_j + \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p} \gamma_{jk} \, lnw_j ln\, w_k + \sum_{j=1}^{p} \gamma_{jQ_i} lnQ_i \, lnw_j$$

Dimana

$\gamma_{ij=}\gamma_{ji}$.

Q_i= Output (Hasil tangkap ikan) Usaha Rumah Tangga

C_i = Total biaya unit usaha ke-i

w_j = Biaya yang dikeluarkan untuk input ke-j

α_0= intersep, dapat diartikan sebagai logaritma dari biaya tetap.

α_Q= parameter yang berhubungan dengan output (hasil tangkapan).

α_j =parameter yang berhubungan dengan input.

γ_{jQ} = parameter yang berhubungan dengan kombinasi output dan input.

γ_{jk}= parameter yang berhubungan dengan kombinasi input.

γ_{QQ}= parameter yang berhubungan dengan output kuadrat.

Jika fungsi translog biaya di atas diturunkan terhadap hasil tangkapan, maka akan didapatkan elastisitas biaya, yang dinyatakan dengan:

$$\frac{\delta \ln TOTALCOST_i}{\delta \ln TOTALHASIL_i} = \alpha_Q + \gamma_{QQ} lnTOTALHASIL_i + \gamma_{1Q} lnTOTALHASIL_i \ln C\_OPERASIONAL_i + \gamma_{2Q} lnTOTALHASIL_i \ln C\_KAPITAL_i + \gamma_{3Q} lnTOTALHASIL_i \ln C\_UPAH_i$$

Dari persamaan di atas, γ_QQmenyatakan perubahan pada elastisitas ketika output berubah. Pada

dasarnya, elastisitas biaya merupakan rasio antara biaya marjinal dengan biaya rata-rata.

$$\frac{\delta \ln TOTALCOST_i}{\delta \ln TOTALHASIL_i} = \frac{\delta TOTALCOST_i}{\delta TOTALHASIL_i}\frac{TOTALHASIL}{TOTALCOST} = \frac{MC}{AC}$$

Indeks skala ekonomis dapat dinyatakan dengan mudah dalam bentuk:

$$SCE = 1 - \frac{\delta \ln TOTALCOST_i}{\delta \ln TOTALHASIL_i}$$

Dari penurunan fungsi biaya terhadap produksi, didapatkan bahwa rumah tangga yang melakukan usaha budidaya bandeng pada tahun 2014 di Sulawesi Selatan yang terjadi skala ekonomis sebesar 70%, jadi persentase rumah tangga yang dapat melipatgandakan output dengan biaya kurang dari dua kali lipat bila input digandakan secara proporsional sebesar 70%. Pada skala dis-ekonomis sebesar 30%, usaha rumah tangga budidaya untuk menggandakan output sebesar dua kali lipat, dibutuhkan biaya input lebih dari dua kali lipat sebesar 30% rumah tangga.

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis data dan pembahasan, serta sesuai dengan tujuan penelitian dapat disimpulkan beberapa hal sebagai berikut:

1. Jumlah terbesar rumah tangga usaha budidaya ikan menurut tempat budidaya di Sulawesi Selatan pada tahun 2013 di tambak/air payau dan kedua di laut.
2. Jumlah terbesar rumah tangga usaha budidaya ikan menurut jenis ikan utama tanpa ikan hias di provinsi Sulawesi Selatan pada tahun 2013 yang pertama melakukan usaha budidaya bandeng dan kedua melakukan usaha budidaya rumput laut.
3. Variabel luas lahan, biaya benih/bibit, biaya pupuk dan obat-obatan, dan biaya tenaga kerja memberikan pengaruh signifikan dan positif terhadap produksi bandeng, variabel yang berpengaruh paling tinggi dan paling rendah terhadap produksi bandeng yaitu tenaga kerja dan luas lahan.
4. Kondisi Return to Scale atau skala pengembalian pada usaha rumah tangga budidaya ikan bandeng di Sulawesi Selatan tahun 2013 berada pada kondisi Increasing Return to Scale.
5. Persentase rumah tangga budidaya ikan bandeng di Sulawesi Selatan tahun 2013 yang berada pada kondisi skala ekonomis (Economic Scale) lebih besar daripada skala dis-ekonomis (Dis-economic Scale).

Berdasarkan hasil kesimpulan penelitian, penulis mengemukaan beberapa saran sebagai berikut:

1. Untuk rumah tangga usaha budidaya ikan bandeng di Sulawesi Selatan, agar meningkatkan pengelolaan luas lahan, biaya benih, biaya pupuk dan obat-obatan, dan biaya tenaga kerja yang berpengaruh signifikan, sehingga produksi ikan bandeng dapat ditingkatkan lebih tinggi.
2. Untuk pemerintah melalui Kementerian Kelautan Perikanan (KKP) agar melakukan pembinaan terhadap rumah tangga yang melakukan usaha budidaya ikan bandeng, sehingga dapat meningkatkan efisiensi teknis dan memperkecil persentase rumah tangga yang memiliki skala dis-ekonomis.
3. Untuk peneliti selanjutnya, direkomendasikan untuk melakukan penelitian dengan beberapa provinsi yang potensial terhadap budidaya bandeng seperti Jawa Timur, Jawa Tengah dan Jawa Barat. Sehingga dapat menganalisis dan membandingkan kondisi rumah tangga usaha budidaya ikan bandeng di provinsi tersebut.

DAFTAR PUSTAKA

Badan Pusat Statistik Provinsi Sulawesi Selatan. 2013. *Laporan Hasil Sensus Pertanian 2013 (Pencacahan Lengkap)*. Makassar: BPS

Badan Pusat Statistik. 2013. *Laporan Hasil Sensus Pertanian 2013 (Pencacahan Lengkap)*. Jakarta: BPS

Badan Pusat Statistik. 2014. *Nilai PDB Menurut Lapangan Usaha Tahun 2011-2013, Laju Pertumbuhan Tahun 2013*. Diunduh di www.bps.go.id

Christensen, Laurits R. dan William H. Greene. 1976. *Economic of Scale in U.S. Electric Power Generation*. The Journal of Political Economy. 84(4) 655-676

Direktorat Jenderal Perikanan Budidaya. 2013. *Laporan Tahunan Direktorat Produksi Tahun 2013.* Jakarta: KKP

Direktorat Jenderal Perikanan Budidaya. 2011. *Peta Sentra Produksi Perikanan Budidaya Tahun 2010.* Jakarta: KKP

Food and Agriculture Organization of the United Nations. 2014. *The State of World Fisheries and Aquaculture*. Rome: FAO

Heryansyah., Said Muhammad dan Sofyan S. 2013. *Analisis Faktor-Faktor yang Mempengaruhi Produksi Nelayan di Kabupaten Aceh Timur*. Jurnal Ilmu Ekonomi. 1(2): 9-15

Kementerian Kelautan dan Perikanan. 2013. *Jumlah Produksi Perikanan Budidaya Tambak Menurut Jenis Ikan dan Provinsi dari Tahun 2006-2012*. Diunduh di http://statistik.kkp.go.id/index.php/statistik/c/9/0/0/0/0/Statistik-Perikanan-Budidaya-Tambak/?pulau_id=&subentitas_id=56&view_data=2&tahun_start=2006&tahun_to=2012&tahun=2009&filter=Lihat+Data+%C2%BB tanggal 28 Juni 2015

Kementerian Kelautan dan Perikanan. 2015. Komoditas Andalan Indonesia Masuki jajaran Produsen Ikan Dunia. Diunduh di http://www.djpb.kkp.go.id/index.php/arsip/c/258/komoditas-andalan-indonesia-masuki-jajaran-produsen-ikan-terbesar-unia/tanggal 29 Juni 2015

Kementerian Kelautan dan Perikanan. 2013. *Kelautan dan Perikanan Dalam Angka 2013*. Jakarta: KKP

Kementerian Perencanaan Pembangunan Nasional dan JICA. 2014. *Analisis Pencapaian Nilai Tukar Nelayan (NTN)*. Jakarta: BAPPENAS

Kumalasari, Dyah. 2004. *Analisis produksi dan keuntungan usaha tambak bandeng di Kabupaten Cilacap*. Skripsi. Surakarta: Fakultas Ekonomi Universitas Sebelas Maret

Situmorang, Jontor. 2007. *Analisis Produktivitas dengan Menggunakan Fungsi Produksi Cobb-Douglas dalam Menentukan Return to Scale pada PT. Perkebunan Nusantara IV Sawit Langkat*. Skripsi. Medan: Fakultas Teknik Universitas Sumatera Utara

PEMODELAN SEBARAN *TOTAL DISSOLVED SOLID* MENGGUNAKAN METODE *MIXED GEOGRAPHICALLY WEIGHTED REGRESSION*

Dadan Kusnandar[1], Naomi Nessyana Debataraja[2], Siti Fitriani[3]

[1,2,3] Program Studi Statistika, FMIPA Universitas Tanjungpura, Pontianak
e-mail: [1]dkusnand@untan.ac.id

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 Bisquare 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.

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*

PENDAHULUAN

Analisis regresi digunakan untuk membangun suatu model matematis untuk menjelaskan hubungan antara dua variabel atau lebih (Kusnandar, Debataraja, Mara, Satyahadewi, 2019). Metode yang umum digunakan untuk menduga parameter model regresi adalah metode *ordinary least square* (OLS). Koefisien regresi pada umumnya dianggap berlaku global untuk seluruh unit pengamatan. Pada kenyataannya, terkadang kondisi antara lokasi pengamatan satu dengan yang lainnya berbeda. Kondisi yang dipengaruhi oleh aspek spasial atau kondisi geografis suatu wilayah penelitian memungkinkan adanya heterogenitas spasial. Heterogenitas spasial merupakan keadaan dimana variabel independen memberikan respon tidak sama pada lokasi yang berbeda dalam satu wilayah penelitian (Tizona, Goejantoro, Wasono, 2017). Data spasial merupakan data yang memuat informasi mengenai letak geografis suatu daerah dan diperoleh dari hasil pengukuran (Fotheringham, Brunsdon, Charlton, 2002). Salah satu metode yang dapat digunakan untuk mengatasi masalah heterogenitas spasial adalah dengan memanfaatkan model *geographically weighted regression* (GWR).

GWR merupakan perluasan dari model regresi. Setiap parameter diduga pada setiap titik sehingga nilai parameter pada GWR berbeda-beda. Suatu pemodelan yang menggabungkan model regresi global dengan model GWR disebut *mixed geographically weighted regression* (*Mixed* GWR) (Fotheringham, dkk, 2002). Seperti halnya pada model GWR, Pendugaan parameter pada model *Mixed* GWR juga dapat dilakukan dengan menggunakan metode *weighted least square* (WLS). Penelitian dengan menggunakan metode *Mixed* GWR telah dilakukan oleh Yasin (2013) yang menggunakan metode resampling bootstrap untuk menguji hipotesis model *Mixed* GWR.

Beberapa pemodelan kualitas air di Kota Pontianak sudah dilakukan oleh Fikri, Debataraja, Kusnandar (2019) dan Kusnandar, Debataraja, Dewi (2019). Penelitian-penelitian tersebut bertujuan untuk menentukan sebaran spasial kualitas air dengan menggunakan analisis diskriminan. Penelitian lain dilakukan oleh Debataraja, Kusnandar, Imro'ah, Rachmadiar (2019) yang menerapkan metode cokriging untuk menduga jumlah zat padat terlarut pada air di Kota Pontianak. Oleh karena itu, metode *Mixed* GWR juga diaplikasikan dalam penelitian ini untuk memodelkan sebaran *Total Dissolved Solid* (TDS) pada 42 lokasi di Kota Pontianak. TDS merupakan salah satu indikator pencemaran air secara fisik yang menunjukkan kandungan padatan terlarut air yang termasuk didalamnya senyawa-senyawa organik dan anorganik.

METODOLOGI

Regresi Linier

Metode regresi linier merupakan metode yang memodelkan hubungan antara variabel dependen dan variabel independen. Model umum regresi linier dapat ditulis sebagai berikut (Chasco, Gracia, Vicens, 2007):

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i$$

dimana y_i adalah nilai pengamatan variabel dependen ke-*i* dengan $i = 1, 2, \ldots, n$; x_{ik} adalah nilai pengamatan variabel independen ke-*k* pada pengamatan ke-*i* dengan $k = 1, 2, \ldots, p$; β_0 adalah konstanta atau intersep model regresi; β_k adalah parameter regresi variabel independen ke-*k* dan ε_i adalah eror pada pengamatan ke-*i*. Metode yang digunakan untuk menduga parameter model regresi adalah dengan meminimumkan jumlah kuadrat eror atau sering disebut dengan OLS yaitu (Kusnandar, dkk, 2019):

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T \boldsymbol{X} \right)^{-1} \boldsymbol{X}^T \boldsymbol{Y}$$

Geographically Weighted Regression

GWR merupakan pengembangan dari model regresi dimana pendugaan parameter dihitung pada setiap lokasi pengamatan,

sehingga mempunyai nilai parameter regresi yang berbeda-beda atau bersifat lokal. Model GWR dapat ditulis sebagai berikut (Fotheringham, dkk, 2002):

$$y_i = \beta_0(r_i, s_i) + \sum_{k=1}^{p} \beta_k(r_i, s_i) x_{ik} + \varepsilon_i$$

dimana y_i adalah nilai pengamatan variabel dependen ke-*i* dengan $i = 1, 2, \ldots, n$; x_{ik} adalah nilai pengamatan variabel independen ke-*k* pada pengamatan ke-*i*; $\beta_0(r_i, s_i)$ adalah konstanta atau intersep model GWR; $\beta_k(r_i, s_i)$ adalah parameter regresi variabel independen ke-*k* pada lokasi pengamatan ke-*i*; (r_i, s_i) adalah titik koordinat lokasi *i* dan ε_i adalah eror pada pengamatan ke-*i*.

Pendugaan parameter model GWR menggunakan WLS yaitu dengan memberikan pembobot yang berbeda untuk setiap lokasi pengamatan. Penduga parameter model untuk setiap lokasinya adalah (Fotheringham, dkk, 2002):

$$\hat{\boldsymbol{\beta}}(r_i, s_i) = \left(\mathbf{X}^T \mathbf{W}(r_i, s_i) \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}(r_i, s_i) \mathbf{Y}$$

dimana $\boldsymbol{X}$ adalah matriks variabel independen $n \times (p+1)$; $\mathbf{Y}$ adalah matriks variabel dependen berukuran $n \times 1$; $\hat{\boldsymbol{\beta}}(r_i, s_i)$ adalah vektor penduga parameter GWR dan $\mathbf{W}(r_i, s_i)$ adalah matriks diagonal berukuran $n \times n$ yang merupakan matriks pembobot spasial lokasi ke-*i*.

Pemilihan Pembobot

Pembobot pada model GWR sangat penting karena fungsi pembobot memberikan hasil pendugaan parameter yang berbeda untuk setiap lokasi. Besarnya pembobotan untuk setiap lokasi yang berbeda dapat ditentukan salah satunya dengan menggunakan fungsi *Kernel*. Pendugaan parameter untuk pembobotan yang digunakan dalam model GWR adalah menggunakan fungsi *Kernel*. Pembobot yang terbentuk dari fungsi *Kernel* terdiri dari (Apriyani, Yuniarti dan Hayati, 2018):

1. *Fixed Gaussian Kernel*

$$w_{ij}(r_i, s_i) = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right)$$

2. *Fixed Bisquare Kernel*

$$w_{ij}(r_i, s_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2 & , \text{untuk } d_{ij} \le b \\ 0 & , \text{untuk } d_{ij} > b \end{cases} \quad \textit{Fixed}$$

3. *Tricube Kernel*

$$w_{ij}(r_i, s_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^3\right)^3 & , \text{untuk } d_{ij} \le b \\ 0 & , \text{untuk } d_{ij} > b \end{cases}$$

dengan *b* adalah *bandwidth* yang sama digunakan untuk setiap lokasi dan d_{ij} adalah jarak antara titik di lokasi *i* dan lokasi *j* yang didapatkan dari jarak *Euclidean* $\left(d_{ij}\right)^2 = \left(r_i - r_j\right)^2 + \left(s_i - s_j\right)^2$.

Dalam penelitian ini, pemilihan *bandwith* optimum dilakukan dengan metode *cross validation* sebagai berikut (Fotheringham, dkk, 2002):

$$CV = \sum_{i=1}^{n} \left(y_i - \hat{y}_{\neq i}(b)\right)^2$$

dengan $\hat{y}_{\neq i}(b)$ adalah nilai pendugaan y_i dimana pengamatan lokasi (r_i, s_i) dihilangkan dari proses pendugaan dan *n* adalah jumlah sampel.

***Mixed Geographically Weighted Regression (Mixed* GWR)**

Model *Mixed* GWR merupakan gabungan dari model regresi global dengan model GWR (Yasin, 2013). Pada model *Mixed* GWR akan dihasilkan penduga parameter yang sebagian bersifat global dan sebagian yang lain bersifat lokal sesuai dengan lokasi pengamatan (Fotheringham, dkk, 2002). Model *Mixed* GWR dengan *p* variabel independen dan *q* variabel independen yang diantaranya bersifat lokal dengan mengasumsikan bahwa intersep model bersifat lokal. Model *Mixed* GWR dapat dituliskan sebagai berikut (Yasin, 2013):

$$y_i = \beta_0(r_i, s_i) + \sum_{k=1}^{q} \beta_k (r_i, s_i) x_{ik} + \sum_{k=q+1}^{p} \beta_k x_{ik} + \varepsilon_i$$

dengan $i = 1, 2, \ldots, n$; y_i adalah nilai pengamatan variabel dependen ke-*i*; x_{ik} adalah nilai pengamatan variabel independen ke-*k* pada pengamatan ke-*i*; $\beta_0(r_i, s_i)$ adalah nilai intersep model regresi; β_k adalah koefisien regresi variabel independen ke-*k*; (r_i, s_i) adalah titik koordinat (lintang, bujur) lokasi *i* dan ε_i adalah eror ke-*i*. Pendugaan parameter model *Mixed* GWR dapat ditulis sebagai berikut (Yasin, 2013):

$$\hat{\boldsymbol{\beta}}_g = \left[\boldsymbol{X}_g^T (\boldsymbol{I} - \boldsymbol{S}_l)^T (\boldsymbol{I} - \boldsymbol{S}_l) \boldsymbol{X}_g\right]^{-1} \boldsymbol{X}_g^T (\boldsymbol{I} - \boldsymbol{S}_l)^T (\boldsymbol{I} - \boldsymbol{S}_l) \boldsymbol{y} \quad (1)$$

dan

$$\hat{\boldsymbol{\beta}}_l (r_i, s_i) = \left[\boldsymbol{X}_l^T \boldsymbol{W} (r_i, s_i) \boldsymbol{X}_l\right]^{-1} \boldsymbol{X}_l^T \boldsymbol{W} (r_i, s_i) \left(\boldsymbol{y} - \boldsymbol{X}_g \hat{\boldsymbol{\beta}}_g\right) \quad (2)$$

dan bentuk dari matriks $\boldsymbol{S}_l$ yang berukuran $n \times n$ yaitu sebagai berikut:

$$\boldsymbol{S}_l = \begin{bmatrix} \boldsymbol{X}_{l1}^T \left[\boldsymbol{X}_1^T \boldsymbol{W} (r_1, s_1) \boldsymbol{X}_l\right]^{-1} \boldsymbol{X}_l^T \boldsymbol{W} (r_1, s_1) \\ \boldsymbol{X}_{l2}^T \left[\boldsymbol{X}_1^T \boldsymbol{W} (r_2, s_2) \boldsymbol{X}_l\right]^{-1} \boldsymbol{X}_l^T \boldsymbol{W} (r_2, s_2) \\ \vdots \\ \boldsymbol{X}_{\ln}^T \left[\boldsymbol{X}_1^T \boldsymbol{W} (r_n, s_n) \boldsymbol{X}_l\right]^{-1} \boldsymbol{X}_l^T \boldsymbol{W} (r_n, s_n) \end{bmatrix}$$

Pemilihan Model Terbaik

Ada beberapa metode yang digunakan untuk memilih model yang sesuai, salah satunya adalah *akaike information criterion* (AIC) yang didefinisikan (Ispriyanti, Yasin, Warsito, Hoyyi dan Winarso, 2017):

$$AIC_c = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n \left\{ \frac{n + tr(\mathbf{S})}{n - 2 - tr(\mathbf{S})} \right\}$$

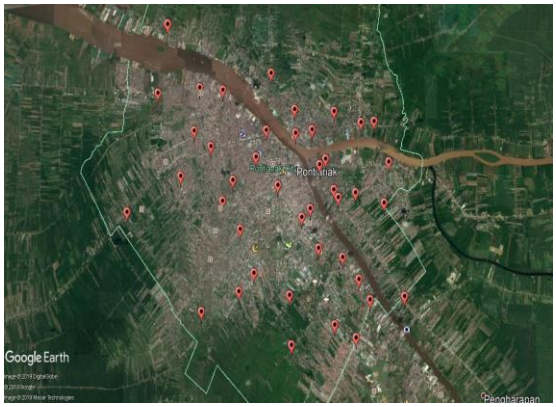
dengan $\hat{\sigma}$ adalah nilai penduga standar deviasi dari eror hasil pendugaan maksimum *likelihood*, yaitu $\hat{\sigma} = \frac{SSE}{n}$ dan $\mathbf{S}$ adalah matriks proyeksi dimana $\hat{y} = \mathbf{S}y$. Pemilihan model terbaik dilakukan dengan memilih nilai AIC terkecil (Fotheringham, dkk, 2002). Berikut merupakan bentuk dari matriks $\mathbf{S}$:

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{X}_1^T \left[\boldsymbol{X}^T \boldsymbol{W} (r_1, s_1) \boldsymbol{X}\right]^{-1} \boldsymbol{X}^T \boldsymbol{W} (r_1, s_1) \\ \boldsymbol{X}_2^T \left[\boldsymbol{X}^T \boldsymbol{W} (r_2, s_2) \boldsymbol{X}\right]^{-1} \boldsymbol{X}^T \boldsymbol{W} (r_2, s_2) \\ \vdots \\ \boldsymbol{X}_{\mathrm{n}}^T \left[\boldsymbol{X}^T \boldsymbol{W} (r_n, s_n) \boldsymbol{X}\right]^{-1} \boldsymbol{X}^T \boldsymbol{W} (r_n, s_n) \end{bmatrix}$$

dengan $\mathbf{S}$ berukuran $n \times n$ dan $\mathbf{W}(r_i, s_i)$ merupakan matriks berukuran $n \times n$.

HASIL DAN PEMBAHASAN

Analisis Data



Gambar 1. Peta Titik Lokasi Sampel

Data yang digunakan dalam penelitian ini merupakan data sampel air yang diambil pada pada 42 lokasi di Kota Pontianak. Pengambilan sampel dilakukan dengan menggunakan metode *stratified random sampling*. Selanjutnya sampel-sampel air tersebut dilakukan analisis di laboratorium. Variabel yang digunakan dalam penelitian ini yaitu *Total Dissolved Solid* (TDS) sebagai variabel dependen (Y) dan variabel independen adalah *Chemical Oxygen Demands* (COD) (X_1), *Biological Oxygen Demands* (BOD) (X_2), pH(X_3), kesadahan (X_4) dan warna (X_5). Titik-titik

lokasi pengambilan sampel air disajikan pada Gambar 1 (Debataraja, dkk, 2019)

Nilai-nilai statistik sampel untuk setiap variabel disajikan dalam Tabel 1. Nilai koefisien keragaman dalam Tabel 1 menunjukkan bahwa data dalam setiap variabel mempunyai keragaman yang relatif besar. Kecuali variabel pH, setiap variabel mempunyai nilai koefisien keragaman di atas 30%. Keragaman yang besar ini juga ditunjukkan oleh nilai kisaran, selisih antara nilai maksimum dan nilai minimum, yang besar.

Pendugaan Model Regresi Global

Pendugaan parameter dalam model regresi global dilakukan dengan menggunakan metode OLS. Persamaan model regresi global yang terbentuk adalah:

$$\hat{y} = 22,33 - 0,19x_1 + 0,53x_2 + 4,94x_3 + 1,02x_4 - 0,04x_5 \quad (3)$$

Nilai *Adjusted R-Square* bagi persamaan tersebut adalah sebesar 79,47% yang berarti bahwa model regresi linier mampu menjelaskan variabel dependen sebesar 79,47% sedangkan 20,53% sisanya dijelaskan oleh variabel lain di luar model.

Tabel 1. Statistik Deskriptif

Variabel	Min	Maks	Rata-rata	Simp. Baku
TDS (mg/l)	17,20	156,30	74,10	37,85
COD (mg/l)	31,05	126,86	81,00	30,91
BOD (mg/l)	7,62	50,50	18,60	8,65
pH	3,88	7,94	6,90	1,05
Kesadahan (mg/l)	8	68	38,80	13,37
Warna (pt.co)	22	990	365,40	329,55

Pengujian signifikansi parameter secara parsial dalam regresi global dapat dilihat pada Tabel 2. Hasil analisis yang tertera pada Tabel 2 dapat diketahui bahwa variabel yang signifikan adalah kesadahan dan warna. Hal ini ditunjukkan dengan nilai *p-value* kecil untuk kedua variabel tersebut (lebih kecil dari α = 5%).

Pengujian asumsi residual dari model regresi global dilakukan untuk menguji

Tabel 2. Uji Signifikansi Parameter Secara Parsial

Kode	Variabel	t_{hitung}	*p-value*
X_1	COD	-1,11	0,27
X_2	BOD	1,44	0,16
X_3	Ph	1,18	0,25
X_4	Kesadahan	3,81	5.10^{-4}
X_5	Warna	-2,46	19.10^{-3}

apakah terdapat pelanggaran terhadap asumsi klasik. Tabel 3 menyajikan hasil uji baik untuk uji multikolinearitas, heterokedastisitas, autokorelasi dan normalitas. Dari Tabel 3, dapat dilihat bahwa semua nilai VIF dari masing-masing variabel (COD, BOD, pH dan kesadahan) kurang dari 10. Hal ini dapat disimpulkan bahwa tidak terjadi multikolinearitas.

Tabel 3 juga menyajikan nilai BP_{hitung}=6,70. Nilai ini kurang dari $\chi^2_{tabel} = 11,07$. Hal ini menunjukkan tidak terjadi heterokedastisitas. Nilai Durbin Watson=1,90 yang terletak diantara dL (1,254) dan 4-dU 1,7814) sehingga dapat disipulkan tidak adanya autokorelasi. Dengan menggunaka $\alpha = 5\%$, nilai *p-value* yang diperoleh sebesar 0,8 sehingga dapat disimpulkan bahwa eror berdistribusi normal.

Tabel 3. Pengujian Asumsi Klasik

Variabel	Nilai VIF
COD	3,81
BOD	1,39
Ph	2,70
Kesadahan	1,79
BP_{hitung}**=6,70**	
DW=1,90	
p-value=0,89	

Dalam model GWR, fungsi pembobot yang digunakan yaitu *Fixed Gaussian Kernel, Fixed Bisquare Kernel* dan *Fixed Tricube Kernel*. Hasil perbandingan dari ketiga fungsi *Kernel* disajikan pada Tabel 3.

Hasil analisis yang disajikan pada Tabel 4 menunjukkan bahwa nilai CV minimum terdapat pada fungsi pembobot *Fixed Bisquare Kernel*. Hal ini berarti bahwa fungsi pembobot *Fixed Bisquare*

Tabel 4. Cross Validation dan Bandwidth Pada Fungsi Pembobot

Fungsi Pembobot	CV minimum	*Bandwidth*
Fixed Gaussian Kernel	12810,11	0,036
Fixed Bisquare Kernel	12685,25	0,091
Fixed Tricube Kernel	12812,73	0,094

Kernel yang digunakan untuk menentukan pembobotan pada model GWR.

Setelah mendapatkan *bandwidth* optimum, langkah selanjutnya adalah menentukan matriks pembobot pada masing-masing lokasi ke-*i* $\mathbf{W}(r_i, s_i)$ dengan menghitung jarak *Euclidean* (r_i, s_i) terlebih dahulu pada setiap lokasi pengamatan. Matriks pembobot digunakan untuk mencari penduga param Neter model GWR dengan memasukkan nilai pembobot kedalam perhitungannya. Setiap nilai diduga pada titik pengamatan, sehingga setiap titik pengamatan mempunyai nilai parameter yang berbeda-beda.

Berdasarkan pendugaan parameter pada regresi global dengan menggunakan metode OLS dapat diperoleh parameter yang berpengaruh secara global $(\hat{\beta}_g)$ yaitu parameter yang signifikan. Selain parameter global $(\hat{\beta}_g)$, parameter yang tidak signifikan diduga memiliki pengaruh secara lokal $\hat{\beta}_l(r_i, s_i)$. Oleh karena itu, parameter yang memiliki pengaruh global dan lokal selanjutnya dapat dibentuk model *Mixed* GWR. Variabel yang berpengaruh secara global yaitu kesadahan (X_4) dan warna (X_5), sedangkan variabel yang berpengaruh secara lokal yaitu COD (X_1), BOD (X_2) dan pH (X_3).

Nilai pendugaan parameter $\hat{\beta}_g$ pada model *Mixed* GWR didapatkan dari perhitungan menggunakan OLS yaitu terdapat pada Persamaan 1. Kemudian, untuk menghitung nilai pendugaan $\hat{\beta}_l(r_i, s_i)$ lokal pada model *Mixed* GWR dihitung berdasarkan Persamaan 2 dengan menggunakan pendugaan WLS.

Pengolahan *Mixed* GWR dilakukan menggunakan *Software R*. Pembobot yang digunakan dalam pemodelan *Mixed* GWR sama seperti model GWR yaitu menggunakan *Fixed Bisquare Kernel*.

Pemilihan Model Terbaik

Setelah menduga parameter model *Mixed* GWR, selanjutnya mencari nilai $\hat{y}$ yang dibandingkan dengan *y* aktualnya. Kemudian perbandingan antara model regresi global, GWR dan *Mixed* GWR dengan menggunakan pembobot *Fixed Bisquare Kernel* dilakukan dengan menggunakan *mean absolute percentage error* (MAPE) dan AIC. Hasil perbandingan kesesuaian model disajikan pada Tabel 5.

Tabel 5. Perbandingan Kesesuaian Model

Model	MAPE	AIC
Regresi Global	26,44%	365,46
GWR	22,52%	346,45
Mixed GWR	22,34%	326,48

Hasil analisis yang tertera pada Tabel 5 dapat dilihat bahwa nilai MAPE terkecil terdapat pada model *Mixed* GWR yang berarti bahwa model *Mixed* GWR memiliki kriteria cukup baik dan nilai AIC untuk melihat model terbaik terdapat pada model *Mixed* GWR. Hasil estimasi parameter mixed GWR untuk 42 lokasi disajikan pada Tabel 6.

Tabel 6. Hasil Estimasi Parameter *Model Mixed GWR*

Lokasi	Model *Mixed* GWR
1	$\hat{y}_1 = -31,78 - 0,18x_1 + 1,06x_2 + 8,56x_3 + 1,30x_4 - 0,04x_5$
2	$\hat{y}_2 = -16,57 - 0,20x_1 + 0,93x_2 + 7,09x_3 + 1,30x_4 - 0,04x_5$
3	$\hat{y}_3 = -34,11 - 0,17x_1 + 1,09x_2 + 8,74x_3 + 1,30x_4 - 0,04x_5$
4	$\hat{y}_4 = -27,06 - 0,18x_1 + 1,00x_2 + 8,13x_3 + 1,30x_4 - 0,04x_5$
5	$\hat{y}_5 = -7,99 - 0,21x_1 + 0,87x_2 + 6,18x_3 + 1,30x_4 - 0,04x_5$

6	$\hat{y}_6 = -0,84 - 0,22x_1 + 0,82x_2 + 5,21x_3 + 1,30x_4 - 0,04x_5$
7	$\hat{y}_7 = -4,36 - 0,22x_1 + 0,81x_2 + 4,81x_3 + 1,30x_4 - 0,04x_5$
8	$\hat{y}_8 = -14,66 - 0,20x_1 + 0,91x_2 + 6,89x_3 + 1,30x_4 - 0,04x_5$
9	$\hat{y}_9 = -13,12 - 0,20x_1 + 0,91x_2 + 6,72x_3 + 1,30x_4 - 0,04x_5$
10	$\hat{y}_{10} = -10,34 - 0,21x_1 + 0,89x_2 + 6,43x_3 + 1,30x_4 - 0,04x_5$
11	$\hat{y}_{11} = 45,57 - 0,18x_1 + 0,52x_2 - 0,18x_3 + 1,30x_4 - 0,04x_5$
12	$\hat{y}_{12} = 17,29 - 0,23x_1 + 0,74x_2 + 3,34x_3 + 1,30x_4 - 0,04x_5$
13	$\hat{y}_{13} = 28,82 - 0,23x_1 + 0,69x_2 + 1,98x_3 + 1,30x_4 - 0,04x_5$
14	$\hat{y}_{14} = -5,46 - 0,21x_1 + 0,86x_2 + 5,90x_3 + 1,30x_4 - 0,04x_5$
15	$\hat{y}_{15} = 11,30 - 0,23x_1 + 0,77x_2 + 4,03x_3 + 1,30x_4 - 0,04x_5$
16	$\hat{y}_{16} = 21,00 - 0,23x_1 + 0,72x_2 + 2,90x_3 + 1,30x_4 - 0,04x_5$
17	$\hat{y}_{17} = 37,07 - 0,22x_1 + 0,63x_2 + 0,96x_3 + 1,30x_4 - 0,04x_5$
18	$\hat{y}_{18} = 27,60 - 0,23x_1 + 0,69x_2 + 2,12x_3 + 1,30x_4 - 0,04x_5$
19	$\hat{y}_{19} = 40,82 - 0,21x_1 + 0,60x_2 + 0,48x_3 + 1,30x_4 - 0,04x_5$
20	$\hat{y}_{20} = 43,60 - 0,17x_1 + 0,51x_2 - 0,19x_3 + 1,30x_4 - 0,04x_5$
21	$\hat{y}_{21} = 26,05 - 0,23x_1 + 0,69x_2 + 2,31x_3 + 1,30x_4 - 0,04x_5$
22	$\hat{y}_{22} = 15,55 - 0,23x_1 + 0,75x_2 + 3,54x_3 + 1,30x_4 - 0,04x_5$
23	$\hat{y}_{23} = 18,52 - 0,23x_1 + 0,73x_2 + 3,20x_3 + 1,30x_4 - 0,04x_5$
24	$\hat{y}_{24} = 45,59 - 0,18x_1 + 0,52x_2 - 0,18x_3 + 1,30x_4 - 0,04x_5$
25	$\hat{y}_{25} = 44,73 - 0,15x_1 + 0,47x_2 - 0,12x_3 + 1,30x_4 - 0,04x_5$
26	$\hat{y}_{26} = 44,36 - 0,19x_1 + 0,56x_2 + 0,01x_3 + 1,30x_4 - 0,04x_5$
27	$\hat{y}_{27} = 44,59 - 0,12x_1 + 0,36x_2 + 0,17x_3 + 1,30x_4 - 0,04x_5$
28	$\hat{y}_{28} = 40,96 - 0,21x_1 + 0,60x_2 + 0,47x_3 + 1,30x_4 - 0,04x_5$
29	$\hat{y}_{29} = 44,10 - 0,20x_1 + 0,56x_2 + 0,05x_3 + 1,30x_4 - 0,04x_5$
30	$\hat{y}_{30} = 42,56 - 0,20x_1 + 0,58x_2 + 0,26x_3 + 1,30x_4 - 0,04x_5$
31	$\hat{y}_{31} = 42,34 - 0,21x_1 + 0,58x_2 + 0,29x_3 + 1,30x_4 - 0,04x_5$
32	$\hat{y}_{32} = 33,78 - 0,22x_1 + 0,65x_2 + 0,37x_3 + 1,30x_4 - 0,04x_5$
33	$\hat{y}_{33} = 32,20 - 0,23x_1 + 0,66x_2 + 1,57x_3 + 1,30x_4 - 0,04x_5$
34	$\hat{y}_{34} = 24,68 - 0,23x_1 + 0,70x_2 + 2,47x_3 + 1,30x_4 - 0,04x_5$
35	$\hat{y}_{35} = 22,19 - 0,23x_1 + 0,71x_2 + 2,77x_3 + 1,30x_4 - 0,04x_5$
36	$\hat{y}_{36} = 23,71 - 0,23x_1 + 0,71x_2 + 2,59x_3 + 1,30x_4 - 0,04x_5$
37	$\hat{y}_{37} = 32,96 - 0,22x_1 + 0,65x_2 + 1,47x_3 + 1,30x_4 - 0,04x_5$
38	$\hat{y}_{38} = 20,39 - 0,23x_1 + 0,72x_2 + 2,98x_3 + 1,30x_4 - 0,04x_5$
39	$\hat{y}_{39} = -15,13 - 0,20x_1 + 0,92x_2 + 6,94x_3 + 1,30x_4 - 0,04x_5$
40	$\hat{y}_{40} = 40,69 - 0,21x_1 + 0,60x_2 + 0,50x_3 + 1,30x_4 - 0,04x_5$
41	$\hat{y}_{41} = -21,69 - 0,19x_1 + 0,96_2 + 7,61x_3 + 1,30x_4 - 0,04x_5$
42	$\hat{y}_{42} = 34,31 - 0,22x_1 + 0,65x_2 + 1,31x_3 + 1,30x_4 - 0,04x_5$

Hasil analisis dengan model *Mixed* GWR menunjukkan bahwa terdapat faktor-faktor yang mempengaruhi TDS di Kota Pontianak yang memiliki pengaruh secara global pada seluruh lokasi pengamatan di Kota Pontianak dan berpengaruh secara lokal pada setiap lokasi pengamatan di Kota Pontianak. Variabel yang berpengaruh secara global yaitu kesadahan dan warna, sedangkan variabel yang berpengaruh secara lokal yaitu COD, BOD dan pH.

KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan, diperoleh kesimpulan bahwa model *Mixed* GWR lebih baik dibandingkan dengan regresi global dan GWR dalam memodelkan sebaran TDS di Kota Pontianak.

DAFTAR PUSTAKA

Apriyani, N.F., Yuniarti, D., dan Hayati, M. N. Pemodelan Mixed Geographically Weighted Regression (Studi Kasus: Jumlah Penderita Diare di Provinsi Kalimantan Timur Tahun 2015. *Jurnal Eksponensial*. 2018. **9**(1). ISSN: 2085-7829.

Chasco, C., Gracia, I., dan Vicens, J. Modeling Spatial Variations in Household Disposable Income with Geographically Weighted Regression. *Munich Personal RePEc Achive Paper*. 2007 (1682).

Debataraja, N. N., Kusnandar, D., Imro'ah, N., dan Rachmadiar, M. Penerapan Metode Cokriging untuk Mengestimasi Jumlah Zat Padat Terlarut Pada Air di Permukiman Kota Pontianak. *Jurnal Matematika, Sains dan Teknologi*. 2019. **20**(2). 142-148.

Debataraja, N. N., Kusnandar, D., dan Nusantara, R. W. Identifikasi Lokasi Sebaran Pencemaran Air di Kawasan Permukiman Kota Pontianak. *Jurnal Matematika, Statistika dan Komputasi*. 2018. **15**(1), 37-41.

Fikri, M., D., Debataraja, N. N., dan Kusnandar, D. Penentuan Sebaran Spasial Pencemaran Air di Kota Pontianak Menggunakan Analisis Diskriminan Dua Kelompok. *Jurnal Media Statistika*. 2019. **12**(2), 226-235.

Fotheringham, A. S., Brunsdon, C., dan Charlton, M. 2002. *Geographically Weighted Regression*. West Sussex: John Wiley and Sons.

Ispriyanti, D., Yasin, H., Warsito, B., Hoyyi, A., dan Winarso, K. Mixed Geographically Weighted Regression Using Adaptive Bandwidth to Modeling of Air Polluter Standard Index. *ARPN Journal of Engineering and Applied Sciences*. 2017. **12**(15).

Kusnandar, D., Debataraja, N.N., Mara, M.N., Satyahadewi, N. 2019. *Metode Statistika dan Aplikasinya*. Pontianak: Untan Press.

Kusnandar, D., Debataraja, N.N., dan Dewi, P. R. Classification of Water Quality in Pontianak City Using Multivariate Statistical Techniques. *Applied Mathematical Sciences*. 2019. **13**(22), 1069-1075

Tizona, A. R., Goejantoro, R., dan Wasono. Pemodelan *Geographically Weighted Regression* (GWR) dengan Fungsi Pembobotan *Adaptive Bisquare Kernel* untuk Angka Kesakitan Demam Berdarah di Kalimantan Timur Tahun 2015. *Jurnal Eksponensial*. 2017. **8**(1), 87-94.

Yasin, H. Uji Hipotesis *Mixed Geographically Weighted Regression* (*Mixed* GWR) dengan Metode *Bootstrap*. *Prosiding Seminar Nasional Statistika. Universitas Diponegoro.* 2013. 527-536.

PENERAPAN PETA KENDALI ATRIBUT KLASIK DAN PETA KENDALI NP BAYES PADA PRODUK CACAT AIR MINUM ASRI DI CV. MULTI REJEKI SELARAS PAYAKUMBUH

Ferra Yanuar[1], Mutiara Fara Nabilla[2], Izzati Rahmi[3]

[1,2,3]Universitas Andalas
e-mail: [1]ferrayanuar@sci.unand.ac.id

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)

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)*

PENDAHULUAN

Air merupakan sumber kehidupan bagi manusia dan makhluk hidup lainnya. Salah satu kegunaan air bagi manusia yaitu air minum, sebagai sumber energi utama bagi tubuh. Semakin berkembangnya teknologi maka semakin berkembang pula penyediaan air minum, salah satunya air minum dalam kemasan. Air minum dalam kemasan (AMDK) adalah air yang diolah dengan menggunakan teknologi tertentu, kemudian dikemas dengan berbagai variasi kemasan, seperti gelas, botol, galon, dan kemasan lainnya.

Perusahaan CV. Multi Rejeki Selaras merupakan salah satu perusahaan yang memproduksi air minum dalam kemasan di kota Payakumbuh, Sumatera Barat. Hasil produksi CV. Multi Rejeki Selaras yaitu produk merek Asri. Produk yang dihasilkan oleh perusahaan ini tidak seluruhnya baik, selalu saja ada produk yang mengalami kecacatan (Andriani, 2016). 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 pengendalian kualitas secara statistik atau tidak adalah peta kendali (*control chart*). Secara umum, terdapat dua kategori dalam peta kendali, yaitu peta kendali variabel dan peta kendali atribut. Pada penelitian ini pengidentifikasian dilakukan terhadap cacat produksi air minum Asri di CV. Multi Rejeki Selaras, sehingga peta kendali yang sesuai untuk kasus ini adalah peta kendali atribut. Pada dasarnya, peta kendali atribut dikembangkan berdasarkan dua macam distribusi, yaitu distribusi Poisson dan distribusi Binomial. Peta kendali atribut yang telah dikembangkan berdasarkan distribusi Binomial diantaranya adalah peta kendali *p* (proporsi ketidaksesuaian) dan peta kendali *np* (banyaknya ketidak-sesuaian) (Montgomery, 2012).

Beberapa metode pendugaan titik yang digunakan untuk menduga parameter diantaranya adalah metode Momen, metode *Maksimum Likelihood Estimation* (MLE), dan metode Bayes (Aunali *et al.*, 2017; Ferra, *et al.* 2013). Metode Bayes merupakan metode pendugaan yang menggabungkan distribusi prior dan distribusi sampel. Distribusi prior merupakan distribusi awal yang memberikan informasi tentang parameter, sehingga bila distribusi prior sudah dapat ditentukan, selanjutnya dapat ditentukan distribusi posterior dengan mengkombi-nasikan informasi dalam distribusi prior dengan informasi data sampel melalui teorema Bayes (Erto & Pallotta, 2015; Ferra dkk, 2020).

Peneliti sebelumnya yaitu Andriani (2016) menggunakan peta kendali multivariat *np* untuk menentukan karakte-ristik kecacatan yang mempunyai kons-tribusi terbesar menyebabkan proses tidak terkendali pada proses produksi air minum Asri di CV. Multi Rejeki Selaras. Res-malani (2020) menggunakan metode Bayes untuk menduga parameter Binomial se-hingga didapatkan batas-batas pengendali dari peta kendali *p* Bayes. Pada penelitian ini akan dilakukan konstruksi peta kendali *np* Bayes untuk mengontrol proses produksi air minum Asri tersebut. Selanjutnya akan dilakukan komparasi kinerja antara peta kendali *np* klasik dengan peta kendali *np* Bayes berdasarkan nilai *Average Run Length* (ARL) terkecil (Montgomery, 2012).

METODE

Pengendalian Kualitas Statistik

Pengendalian kualitas statistik merupakan penggunaan metode statistika untuk mengumpulkan dan menganalisa data dalam menentukan dan mengawasi kualitas hasil produk. Salah satu alat bantu untuk menentukan apakah proses dalam keadaan terkendali atau tidak adalah peta kendali. Peta kendali yang dikembangkan oleh Walter Shewhart dikenal dengan sebutan peta kendali Shewhart. Pada peta kendali biasanya dipilih batas toleransi sebesar tiga

kali standar deviasi agar memperoleh kepercayaan sebesar 99,73% untuk mengetahui apakah suatu produksi berada dalam batas kendali atau tidak.

Peta Kendali *np* (Montgomery, 2012)

Peta kendali *np* digunakan untuk mengukur jumlah produk yang tidak sesuai dalam suatu subgrup *k* dengan ukuran sampel produk dalam subgrup setiap kali pengamatan berukuran *n* konstan. Peta kendali ini berdasarkan pada sebaran Binomial dan dikembangkan berdasarkan peta kendali Shewhart sehingga diperlukan nilai harapan dan standar deviasi dari *np* untuk membentuknya. Asas statistik yang melandasi pada peta kendali *np* yaitu distribusi Binomial. Adapun nilai harapan dan ragam dari peta kendali *np* masing-masing adalah:

$$\mu_{n\hat{P}} = nE(\hat{P}) = np,$$
$$\sigma^2_{n\hat{P}} = n^2 Var(\hat{P}) = np(1-p),$$

dengan $\hat{P} = \frac{R}{n}$ didefinisikan sebagai rasio banyaknya produk yang tidak sesuai dalam sampel R dengan ukuran sampel *n* yang merupakan perkiraan nilai yang tidak diketahui dari parameter *p*.

Untuk mengestimasi nilai *p* yang tidak diketahui dari data, digunakan rata-rata dari keseluruhan proporsi yaitu $\bar{p}$ pada ukuran sampel *n* konstan dengan banyaknya pengamatan *k* dinyatakan oleh:

$$\bar{p} = \frac{1}{k}\sum_{i=1}^{k} \hat{p}_i = \frac{\sum_{i=1}^{k} r_i}{nk},$$

sehingga diperoleh batas-batas pengendali peta kendali *np* sebagai berikut:

$$CL = \mu_{n\hat{p}} = n\bar{p}$$
$$UCL = \mu_{n\hat{p}} + \sigma^2_{n\hat{p}} = n\bar{p} + n\bar{p}(1-\bar{p})$$
$$LCL = \mu_{n\hat{p}} - \sigma^2_{n\hat{p}} = n\bar{p} - n\bar{p}(1-\bar{p}).$$

Average Run Length (ARL)

Kriteria yang digunakan untuk dapat membandingkan kinerja peta kendali adalah dengan mengukur seberapa cepat peta kendali tersebut menangkap sinyal *out of control*. Peta kendali yang lebih cepat mendeteksi sinyal *out of control* disebut lebih sensitif terhadap perubahan proses. Salah satu cara untuk mengukur kinerja peta kendali adalah dengan menggunakan *Average Run Length* (ARL).

ARL dinyatakan sebagai :

$$ARL = \frac{1}{P(suatu\ titik\ di\ luar\ kendali)}$$

atau

$$ARL_0 = \frac{1}{\alpha}$$

bila proses dalam keadaan terkendali (i*n control*). Parameter α menyatakan probabilitas kesalahan (*error*) tipe I (menyatakan keadaan tidak terkendali padahal keadaan terkendali) atau probabilitas suatu titik rata-rata sampel jatuh di luar batas pengendali pada saat proses terkendali. Sedangkan perumusan untuk proses dalam keadaan tidak terkendali (*out of control*) adalah:

$$ARL_1 = \frac{1}{1-\beta}$$

dimana β = probabilitas kesalahan (*error*) tipe II (menyatakan keadaan terkendali padahal keadaan tidak terkendali) atau probabilitas suatu titik rata-rata sampel jatuh di dalam batas pengendali pada saat proses tidak terkendali. Probabilitas kesalahan (*error*) tipe II dari peta kendali bagian ketidaksesuaian dapat dihitung dari:

$$\beta = P(\hat{p} < UCL|p) - P(\hat{p} \leq LCL|p).$$

Dengan adanya kedua jenis ARL ini, maka peta kendali terbaik dapat dipilih. Untuk ARL_0 peta kendali terbaik adalah peta kendali dengan nilai ARL terbesar, sedangkan untuk ARL_1 peta kendali terbaik adalah peta kendali dengan nilai ARL terkecil.

Sumber Data

Data kasus yang digunakan adalah data cacat keseluruhan produksi air minum Asri dalam kemasan 240 ml di CV. Multi Rejeki Selaras Kota Payakumbuh pada bulan Juni 2016. Data kasus yang digunakan dalam penelitian ini diambil oleh karyawan bagian produksi setiap hari, yaitu sebanyak 28 hari pada bulan Juni (Andriani, 2016). Pada data ini, banyaknya sampel perhari konstan yaitu 30000. Dengan demikian dinotasikan banyak pengamatan *k* = 28 dan ukuran sampel tiap pengamatan *n* = 30000. Banyaknya produk yang cacat perhari (r_i), $i = 1, 2, \ldots, k$.

Variabel Penelitian

Variabel yang digunakan adalah data produksi air minum Asri dan banyaknya produk yang tidak sesuai dalam proses produksi air minum Asri di CV. Multi Rejeki Selaras Kota Payakumbuh.

Metode Analisis Data

Metode yang digunakan adalah metode peta kendali *np* Bayes dengan melakukan pendugaan titik bagi parameter *p* pada distribusi Binomial dengan menggunakan metode Bayes dan dilanjutkan dengan menentukan batas-batas pengendali pada peta kendali *np* Bayes. Adapun langkah-langkahnya sebagai berikut:

1. Menentukan fungsi likelihood dari distribusi Binomial(1,p).
2. Menentukan distribusi prior konjugat dan distribusi prior non-informatif dari *p*.
3. Menentukan distribusi posterior dari masing-masing distribusi prior yang telah diperoleh.
4. Menentukan UCL, CL, dan LCL dari peta kendali *np* Bayes berdasarkan nilai harapan dan ragam $n\hat{p}$ Bayes dari masing-masing distribusi posterior yang telah diperoleh.

Selanjutnya mengaplikasikan dan membandingkan peta kendali *np* klasik dan peta kendali *np* Bayes pada data kasus. Adapun langkah-langkah pembuatan kedua peta kendali tersebut adalah sebagai berikut:

1. Menggunakan data yang diambil dari CV. Multi Rejeki Selaras Kota Payakumbuh dan menentukan batas pengendalinya untuk peta kendali *np* klasik dan peta kendali *np* Bayes.
2. Menentukan banyaknya sampel yang tidak terkendali pada peta kendali *np* klasik dan peta kendali *np* Bayes berdasarkan data yang diambil dari CV. Multi Rejeki Selaras Kota Payakumbuh.
3. Mencari nilai ARL untuk peta kendali *np* klasik dan peta kendali *np* Bayes lalu membandingkannya.
4. Menentukan peta kendali terbaik berdasarkan nilai ARL terkecil.

HASIL DAN PEMBAHASAN

1. Konstruksi Peta Kendali *np* Bayes

Pada peta kendali *np* Bayes, batas-batas pengendalinya akan diestimasi dengan menggunakan metode Bayes. Misalkan peubah acak *R* adalah peubah acak diskret yang menyatakan banyaknya produk yang tidak sesuai dalam *k* pengamatan dengan ukuran sampel produk tiap pengamatan berukuran *n* konstan dan *p* adalah nilai kemungkinan dari banyaknya produk yang tidak sesuai dalam distribusi Binomial yang hanya memiliki nilai antara 0 dan 1. Selanjutnya dapat dituliskan $R \sim BIN(n,p)$ dengan fungsi kepekatan peluang bersyarat untuk *r* jika diberikan *p* adalah sebagai berikut:

$$f(r|p) = \binom{n}{r} p^r (1-p)^{n-r}, \qquad 0 \le p \le 1.$$

Distribusi Beta merupakan keluarga eksponensial dari distribusi Binomial, sehingga distribusi Beta dapat digunakan sebagai prior konjugat dari distribusi Binomial. Selain itu, parameter *p* dalam distribusi Beta sebagai penduga yang tidak diketahui merupakan peluang banyaknya produk yang tidak sesuai dalam distribusi Binomial yang hanya memiliki nilai antara 0 sampai 1. Karena itu, digunakannya Beta sebagai distribusi prior dengan meng-asumsikan bahwa parameter *p* dapat menjalani banyaknya tak hingga nilai-nilai bilangan riil dari 0 sampai 1. Misalkan $P \sim Beta(\alpha, \beta)$ dan $g(p)$ adalah fungsi kepekatan peluangnya, maka distribusi posterior untuk konjugat Beta dapat diperoleh dari rasio fungsi kepekatan peluang bersama *r* dan *p* dengan fungsi kepekatan peluang marjinal *r* sebagai berikut (Erto *et al.*, 2019):

$$g(p|r) = \frac{f(r|p)g(p)}{\int_{-\infty}^{\infty} f(r|p)g(p)} = \frac{p^{r+\alpha-1}(1-p)^{n-r+\beta-1}}{B(r+\alpha, n-r+\beta)}, \quad 0 < p < 1.$$

Pendugaan Bayes untuk parameter *p* dapat diperoleh dari nilai harapan distribusi posterior. Batas-batas pengendali peta

kendali *np* Bayes prior konjugat Beta berdasarkan nilai harapan dan ragam $n\hat{p}$ Bayes adalah sebagai berikut:

$$CL = n\frac{(n\bar{p}+\alpha)}{n+\alpha+\beta}$$

$$UCL = n\frac{(n\bar{p}+\alpha)}{n+\alpha+\beta} + 3\sqrt{n^2\frac{n\bar{p}(1-\bar{p})}{(n+\alpha+\beta)^2}}$$

$$LCL = n\frac{(n\bar{p}+\alpha)}{n+\alpha+\beta} - 3\sqrt{n^2\frac{n\bar{p}(1-\bar{p})}{(n+\alpha+\beta)^2}},$$

dengan $\alpha = \bar{p}\left(\frac{\bar{p}(1-\bar{p})}{s^2}-1\right)$ dan $\beta = (1-\bar{p})\left(\frac{\bar{p}(1-\bar{p})}{s^2}-1\right)$ diperoleh dari pendugaan momen terhadap parameter Beta.

Salah satu pemilihan distribusi prior non-informatif adalah distribusi Uniform, karena tidak mengandung informasi tentang parameter *p*. Misalkan peubah acak $P \sim Uniform(0,1)$ dan $g(p)$ adalah fungsi kepekatan peluangnya, maka distribusi posterior untuk prior non-informatif dapat diperoleh dari rasio fungsi kepekatan peluang bersama *r* dan *p* dengan fungsi kepekatan peluang marjinal *r* sebagai berikut:

$$g(p|r) = \frac{f(r|p)g(p)}{\int_{-\infty}^{\infty} f(r|p)g(p)}$$

$$= \frac{p^r(1-p)^{n-r}}{B(r+1, n-r+1)}, \quad 0 < p < 1.$$

Batas-batas pengendali peta kendali *np* Bayes prior non-informatif berdasarkan nilai harapan dan ragam $n\hat{p}$ Bayes adalah sebagai berikut:

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

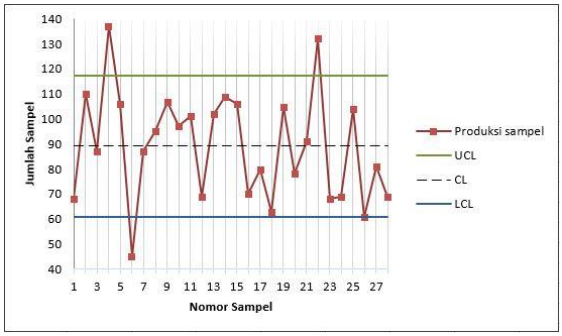
$$UCL = n\frac{(n\bar{p}+1)}{n+2} + 3\sqrt{n^2\frac{n\bar{p}(1-\bar{p})}{(n+2)^2}}$$

$$LCL = n\frac{(n\bar{p}+1)}{n+2} - 3\sqrt{n^2\frac{n\bar{p}(1-\bar{p})}{(n+2)^2}},$$

2. Grafik Peta Kendali *np*

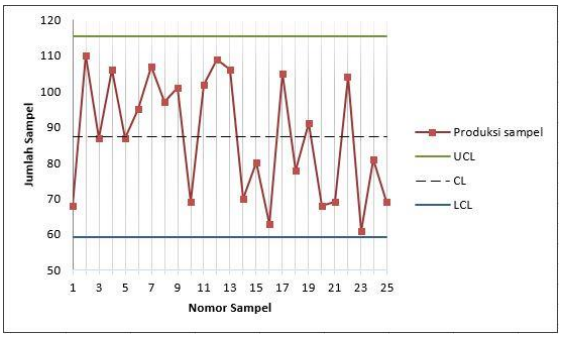
Grafik Peta Kendali *np* Klasik

Gambar 1 memperlihatkan batas-batas pengendali untuk peta kendali *np* klasik. Terlihat bahwa sampel ke 4,6, dan 22 berada di luar batas pengendali, sehingga dapat dikatakan proses produksi dalam keadaan *out of control* (tidak terkendali).



Gambar 1. Peta Klasik *np* Klasik

Selanjutnya dilakukan revisi terhadap peta kendali *np* klasik dengan menghilangkan sampel yang *out of control*, sehingga diperoleh grafik peta kendali *np* klasik (revisi) seperti Gambar 2 berikut.



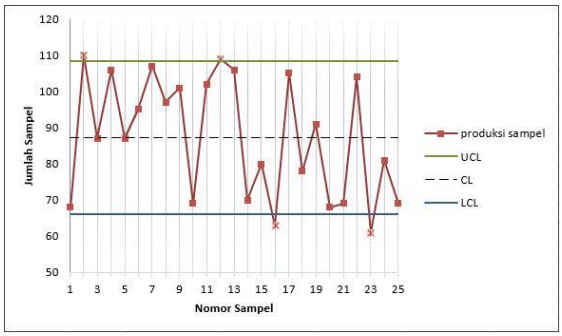
Gambar 2. Peta Kendali *np* Klasik (revisi)

Gambar 2 memperlihatkan bahwa semua sampel berada dalam batas pengendali sehingga proses produksi dalam keadaan *in control* (terkendali).

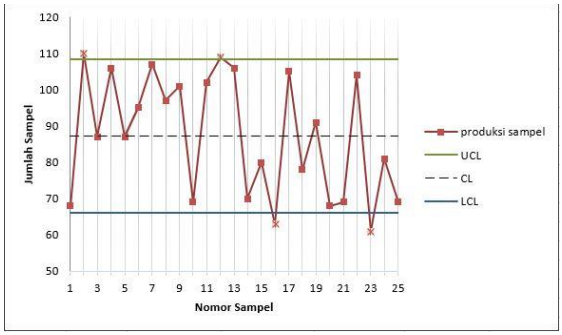
Grafik Peta Kendali *np* Bayes

Gambar 3 memperlihatkan batas-batas pengendali untuk peta kendali *np* Bayes prior konjugat. Dapat dilihat bahwa sampel ke 2, 12, 16, dan 23 berada diluar batas pengendali sehingga proses produksi dalam keadaan *out of control* (tidak terkendali).

Gambar 4 memperlihatkan batas-batas pengendali untuk peta kendali *np* Bayes prior non-informatif. Terlihat bahwa semua sampel berada dalam batas



Gambar 3. Peta Kendali *np* Bayes Prior Konjugat.



Gambar 4. Peta Kendali np Bayes Prior Non-informatif.

Tabel 1. Perbandingan Nilai ARL Peta Kendali np Klasik dan Peta Kendali np Bayes

p	Peta Kendali np Klasik	Peta Kendali np Bayes Konjugat	Peta Kendali np Bayes Non-informatif
0,0019	1,608492842	1,127522832	1,492537
0,0020	2,154708037	1,268874508	1,937984
0,0021	3,097893432	1,52578578	2,725538
0,0022	4,852013586	1,968503937	4,132231
0,0023	8,26446281	2,725538294	6,807352
0,0024	14,97005988	4,07996736	11,93317
0,0025	28,49002849	6,497725796	22,42152
0,0026	58,82352941	11,0619469	43,85965
0,0027	125	19,76284585	93,45794
0,0028	263,1578947	34,12969283	208,3333
0,0028	344,8275862	41,15226337	270,2703
0,0029	370,3703704	43,66812227	344,8276
0,0029	312,5	39,5256917	357,1429
0,0031	147,0588235	24,09638554	188,6792
0,0031	98,03921569	17,76198934	123,4568
0,0031	61,72839506	13,03780965	81,96721
0,0032	40,98360656	9,861932939	53,19149
0,0033	19,8019802	5,827505828	24,44988
0,0034	10,70663812	3,829950211	12,85347
0,0035	6,402048656	2,725538294	7,369197
0,0036	4,132231405	2,082899396	4,719207
0,0037	2,933411558	1,681237391	3,241491
0,0038	2,21141088	1,431639227	2,421894
0,0039	1,77430802	1,273560876	1,908761
0,0040	1,50060024	1,172195522	1,579529

pengendali sehingga proses produksi dalam keadaan *in control* (terkendali).

3. Perbandingan Nilai ARL

Perbandingan nilai ARL untuk peta kendali *np* klasik, peta kendali *np* Bayes prior konjugat dan peta kendali *np* Bayes prior non-informatif dapat dilihat pada Tabel 1.

Dari Tabel 1 teramati bahwa peta kendali *np* Bayes prior konjugat memiliki nilai ARL yang relatif lebih kecil dibandingkan peta kendali *np* klasik dan peta kendali *np* Bayes prior non-informatif untuk setiap nilai *p* yang dipilih. Dengan demikian dapat dikatakan peta kendali yang memiliki kinerja terbaik pada penelitian ini adalah peta kendali *np* Bayes prior konjugat.

KESIMPULAN

Untuk mengkonstruksi peta kendali *np* Bayes sehingga didapatkan batas-batas pengendalinya, terlebih dahulu dilakukan pendugaan terhadap parameter *p* dengan metode Bayes, kemudian menentukan nilai harapan dan ragam bagi $n\hat{p}$ Bayes. Batas-batas pengendali untuk peta kendali *np* Bayes adalah sebagai berikut:

1. Prior Konjugat

$$CL = n\frac{(n\bar{p}+\alpha)}{n+\alpha+\beta}$$

$$UCL = n\frac{(n\bar{p}+\alpha)}{n+\alpha+\beta} + 3\sqrt{n^2\frac{n\bar{p}(1-\bar{p})}{(n+\alpha+\beta)^2}}$$

$$LCL = n\frac{(n\bar{p}+\alpha)}{n+\alpha+\beta} - 3\sqrt{n^2\frac{n\bar{p}(1-\bar{p})}{(n+\alpha+\beta)^2}},$$

2. Prior Non-informatif

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

$$UCL = n\frac{(n\bar{p}+1)}{n+2} + 3\sqrt{n^2\frac{n\bar{p}(1-\bar{p})}{(n+2)^2}}$$

$$LCL = n\frac{(n\bar{p}+1)}{n+2} - 3\sqrt{n^2\frac{n\bar{p}(1-\bar{p})}{(n+2)^2}},$$

Berdasarkan data dari CV. Multi Rejeki Selaras Kota Payakumbuh pada bulan Juni 2016, diperoleh hasil bahwa peta kendali *np* Bayes lebih banyak mengandung data *out of control*, sehingga dapat dikatakan peta kendali *np* Bayes lebih sensitif dan lebih efisien dalam mengontrol data *out of control* dibandingkan peta kendali *np* klasik. Dari perhitungan nilai ARL, peta kendali *np* Bayes prior konjugat memiliki nilai ARL yang relatif lebih kecil dibandingkan peta kendali *np* klasik dan peta kendali *np* Bayes prior non-informatif untuk setiap nilai *p* yang dipilih. Dengan demikian, dapat dikatakan peta kendali yang memiliki kinerja terbaik pada penelitian ini adalah peta kendali *np* Bayes prior konjugat.

Saran yang dapat diberikan untuk penelitian berikutnya agar dapat melakukan analisis peta kendali variabel dengan metode Bayes dan membandingkannya dengan peta kendali variabel yang telah ada sebelumnya.

DAFTAR PUSTAKA

Andriani, Sisi. (2016). Pengontrolan Kualitas Produk Menggunakan Metode Bagan Kendali Multivariat *np* dalam Usaha Peningkatan Kualitas Kasus di CV. Multi Rejeki Selaras Kota Payakumbuh. *Jurnal Matematika*, Vol VI, No. 1, pp: 161-167.

Aunali, AS., Venkatesan D, dan Michele G. (2019). Bayesian Control Charts Using Gamma Prior. *International Journal of Scientific & Technology Research*, Vol. 8, No. 12, pp : 1567 - 1570.

Erto, P. dkk. *(*2019). A Bayesian Control Chart for Monitoring the Ratio of Weibull Percentile. *Quality Reliability Enginering*, Special Issue Article. pp: 1-16.

Erto P dan Pallotta G. (2015). A Semi-empirical Bayeisn Chart to Monitor Weibull Percentile. *Scandinavian Journal of Statistics*. Vol 42, pp: 701-712.

Ferra Y., Kamarulzaman I. dan Abdul AJ. (2013). Bayesian structural equation modeling for the health index. *Journal of Applied Statistics*, Vol. 40, No. 6, pp: 1254-1269.

Ferra Y., Cici S dan Dodi D. (2020). Bayesian Inference for Pareto Distribution with Prior Conjugate and Prior Non-Conjugate. *Jurnal Matematika. Statistika & Komputasi*, No. 16, No. 3, pp: 382 - 390.

Fu, JC., Spiring, FA., dan Xie, H. (2002). On the Average Run Lengths of Quality Control Schemes Using a Markov Chain Approach. *Statistics & Probability Letters*, Vol. 56, pp: 369 - 380.

Kandil, AM., Hamed, MS., dan Mohamed, SM. (2013). Average Run Length for Multivariate T^2 Control Chart Technique with Application. *Journal of the Egyptian Statistical Society*, Vol. 29, No. 2, pp: 1 – 20.

Khoo, MBC., The, SY., Chew, XY., dan Teoh WL. (2015). Standard Deviation of the Run Length (SDRL) and Average Run Length (ARL) Performances of EWMA and Synthetic Charts. *International Journal of Engineering and Technology*, Vol. 7, No. 6, pp: 1 – 4.

Montgomery, D.C. (2012). *Introduction to Statistical Quality Control*, 7th ed., Wiley, New York, NY.

Resmalani, F. (2020). Perbandingan Peta Kendali p Klasik dan Peta Kendali p Bayes serta Aplikasinya pada Data Kasus di PT.XYZ. *Jurnal Matematika*, Vol IX, No. 2, pp: 162-168.

Sunthornwat, R dan Areepong, Y. (2019). Average Run Length on CUSUM Control Chart for Seasonal and Non-Seasonal Moving Average Process with Exogenous Variables. *Symmetry*, Vol 12, No. 173, pp: 1- 15

SMALL AREA ESTIMATION WITH EXCESS ZERO (STUDI KASUS: ANGKA KEMATIAN BAYI DI PULAU JAWA)

Nofita Istiana[1]

[1]Politeknik Statistika STIS
e-mail: [1]nofita@stis.ac.id

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 rataan 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*

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*

PENDAHULUAN

Angka kematian dan kesakitan merupakan indeks kesehatan yang dapat digunakan untuk membandingkan status kesehatan penduduk dari waktu ke waktu atau antar penduduk pada satu titik waktu. Ukuran ini memungkinkan adanya perbandingan sistem dan program kesehatan serta dapat menyoroti penduduk yang membutuhkan perhatian khusus dari layanan kesehatan (Murray *et al*., 2000). Salah satu jenis angka kematian yang penting yaitu angka kematian bayi (AKB).

AKB diperoleh dari Survei Demografi dan Kesehatan Indonesia (SDKI). SDKI merupakan bagian dari program internasional *Demographic and Health Survey* (DHS). Survei ini dirancang untuk mengumpulkan data fertilitas, keluarga berencana, dan kesehatan ibu dan anak. SDKI dilaksanakan bersama oleh Badan Pusat Statistik (BPS), Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN), dan Kementerian Kesehatan setiap lima tahun sekali, dengan panduan dari *United States Agency for International Development* (USAID).

Pendugaan AKB yang dilakukan berdasarkan data SDKI hanya berskala nasional atau provinsi namun dengan adanya sistem desentralisasi diperlukan prediksi untuk area yang lebih kecil seperti level kabupaten, kecamatan maupun kelurahan/desa. Pendugaan untuk level kabupaten, kecamatan maupun kelurahan/desa terkadang masih sulit dilakukan karena memiliki ukuran sampel yang relatif kecil dan terdapat area yang tidak tersampel.

Menurut Rao dan Molina (2015), jika terdapat sub populasi dengan ukuran sampel yang kecil maka pendugaan langsung dapat menghasilkan standard error yang sangat besar. Selain itu, pendugaan langsung tidak dapat dilakukan untuk sub populasi yang tidak memiliki sampel. Oleh karena itu diperlukan teknik pendugaan tidak langsung yang dapat meningkatkan efektivitas ukuran sampel sehingga dapat menurunkan galat baku. Salah satu teknik analisis yang dapat digunakan untuk mengatasi masalah tersebut adalah pendugaan area kecil (*small area estimation*/SAE). SAE dapat dilakukan dengan menggunakan pemodelan yang di dalamnya memasukkan peubah penyerta sebagai pengaruh tetap dan area sebagai pengaruh acak (model campuran). Pemodelan yang digunakan yaitu Model Linier Terampat Campuran (MLTC)/ *Generalized Linear Mixed Model* (GLMM).

Data AKB merupakan data cacahan dengan peluang kejadian sangat kecil, sehingga dapat diasumsikan mengikuti sebaran Poisson. Sebaran Poisson memiliki asumsi rataan sama dengan ragam (equidispersi). Data dengan ragam lebih besar daripada rataan disebut overdispersi, sedangkan data dengan ragam lebih kecil daripada rataan disebut dengan underdispersi. Menurut Dean dan Lundy (2016), terlanggarnya asumsi equidispersi (permasalahan dispersi) dapat disebabkan oleh nilai nol lebih besar dari yang diharapkan (*excess zero*).

Menurut Faraway (2016), permasalahan dispersi pada model akan sangat berpengaruh dalam uji hipotesis untuk parameter dugaan prediksi. Jika model mengalami overdispersi maka simpangan baku statistik menjadi bias ke bawah (*underestimate*), sehingga pengujian hipotesis untuk parameter yang bersesuaian mempunyai kecenderungan untuk menolak H_0 (statistik dianggap signifikan). Sebaliknya, jika model mengalami underdispersi, maka simpangan baku statistik menjadi bias ke atas (*overestimate*), sehingga pengujian hipotesis untuk parameter yang bersesuaian mempunyai kecenderungan untuk tidak menolak H_0 (statistik dianggap tidak signifikan). Permasalahan dispersi biasanya ditangani dengan model *Negative Binomial*. Akan tetapi, model ini tidak banyak memperbaiki permasalahan dispersi pada data cacahan. Hal ini seperti penelitian yang dilakukan oleh Anggreyani (2016) dan Yudistira (2018). Oleh karena itu, penelitian ini menggunakan GLMM dengan *zero inflated* untuk mengatasi permasalahan dispersi yg diakibatkan oleh *excess zero*. Hasil dugaan model ini selanjutnya

dibandingkan dengan hasil dugaan Poisson *mixed model* dan dugaan langsungnya.

METODE

Pendugaan Area Kecil

Angka kematian dan kesakitan merupakan indeks kesehatan yang dapat digunakan untuk membandingkan status kesehatan penduduk dari waktu ke waktu atau antar penduduk pada satu titik waktu. Ukuran ini memungkinkan adanya perbandingan sistem dan program kesehatan serta dapat menyoroti penduduk yang membutuhkan perhatian khusus dari layanan kesehatan (Murray *et al*., 2000). Salah satu jenis angka kematian yang penting yaitu angka kematian bayi (AKB).

AKB diperoleh dari Survei Demografi dan Kesehatan Indonesia (SDKI). SDKI merupakan bagian dari program internasional *Demographic and Health Survey* (DHS). Survei ini dirancang untuk mengumpulkan data fertilitas, keluarga berencana, dan kesehatan ibu dan anak. SDKI dilaksanakan bersama oleh Badan Pusat Statistik (BPS), Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN), dan Kementerian Kesehatan setiap lima tahun sekali, dengan panduan dari *United States Agency for International Development* (USAID).

Pendugaan AKB yang dilakukan berdasarkan data SDKI hanya berskala nasional atau provinsi namun dengan adanya sistem desentralisasi diperlukan prediksi untuk area yang lebih kecil seperti level kabupaten, kecamatan maupun kelurahan/desa. Pendugaan untuk level kabupaten, kecamatan maupun kelurahan/desa terkadang masih sulit dilakukan karena memiliki ukuran sampel yang relatif kecil dan terdapat area yang tidak tersampel.

Menurut Rao dan Molina (2015), jika terdapat sub populasi dengan ukuran sampel yang kecil maka pendugaan langsung dapat menghasilkan standard error yang sangat besar. Selain itu, pendugaan langsung tidak dapat dilakukan untuk sub populasi yang tidak memiliki sampel. Oleh karena itu diperlukan teknik pendugaan tidak langsung yang dapat meningkatkan efektivitas ukuran sampel sehingga dapat menurunkan galat baku. Salah satu teknik analisis yang dapat digunakan untuk mengatasi masalah tersebut adalah pendugaan area kecil (*small area estimation*/SAE). SAE dapat dilakukan dengan menggunakan pemodelan yang di dalamnya memasukkan peubah penyerta sebagai pengaruh tetap dan area sebagai pengaruh acak (model campuran). Pemodelan yang digunakan yaitu Model Linier Terampat Campuran (MLTC)/ *Generalized Linear Mixed Model* (GLMM).

Data AKB merupakan data cacahan dengan peluang kejadian sangat kecil, sehingga dapat diasumsikan mengikuti sebaran Poisson. Sebaran Poisson memiliki asumsi rataan sama dengan ragam (equidispersi). Data dengan ragam lebih besar daripada rataan disebut overdispersi, sedangkan data dengan ragam lebih kecil daripada rataan disebut dengan underdispersi. Menurut Dean dan Lundy (2016), terlanggarnya asumsi equidispersi (permasalahan dispersi) dapat disebabkan oleh nilai nol lebih besar dari yang diharapkan (*excess zero*).

Menurut Faraway (2016), permasalahan dispersi pada model akan sangat berpengaruh dalam uji hipotesis untuk parameter dugaan prediksi. Jika model mengalami overdispersi maka simpangan baku statistik menjadi bias ke bawah (*underestimate*), sehingga pengujian hipotesis untuk parameter yang bersesuaian mempunyai kecenderungan untuk menolak H_0 (statistik dianggap signifikan). Sebaliknya, jika model mengalami underdispersi, maka simpangan baku statistik menjadi bias ke atas (*overestimate*), sehingga pengujian hipotesis untuk parameter yang bersesuaian mempunyai kecenderungan untuk tidak menolak H_0 (statistik dianggap tidak signifikan). Permasalahan dispersi biasanya ditangani dengan model *Negative Binomial*. Akan tetapi, model ini tidak banyak memperbaiki permasalahan dispersi pada data cacahan. Hal ini seperti penelitian yang dilakukan oleh Anggreyani (2016) dan Yudistira (2018). Oleh karena itu, penelitian ini menggunakan GLMM dengan

zero inflated untuk mengatasi permasalahan dispersi yg diakibatkan oleh *excess zero*. Hasil dugaan model ini selanjutnya dibandingkan dengan hasil dugaan Poisson *mixed model* dan dugaan langsungnya.

METODOLOGI

Pendugaan Area Kecil

Pendugaan area kecil merupakan suatu metode untuk menduga parameter pada area kecil dalam percontohan survei dengan memanfaatkan informasi dari luar area, dari dalam area itu sendiri, dan dari luar survei (Kurnia, 2009). Pendugaan area kecil dibagi dua, yaitu *design based* dan *model based* (Rao dan Molina, 2015). *Design based* menggunakan penimbang survei dan desain survei dalam proses inferensianya. Sedangkan *model based* atau pendugaan tidak langsung menggunakan informasi tambahan dalam proses inferensianya. Pendugaan tidak langsung dilakukan dengan cara menambahkan informasi dari area sekitarnya *(neighbouring areas*) dan peubah-peubah penyerta yang berasal dari sensus atau catatan administrasi dan memiliki hubungan dengan peubah yang diamati sehingga dapat meningkatkan keefektifan ukuran sampel. Pendugaan tidak langsung dalam penelitian ini dilakukan dengan menggunakan GLMM.

Excess Zero

Model Poisson digunakan untuk memodelkan data cacahan. Model ini juga memperbolehkan peubah responnya berupa *zero response*. Peluang nilai nol yang dapat ditolerir untuk sebaran Poisson yaitu (Winkleman, 2008)

$$f_0 = e^{-\lambda}. \qquad (1)$$

Suatu penelitian dapat dijumpai kondisi di mana terlalu banyak nol ($f_0 > e^{-\lambda}$) atau terlalu sedikit nol ($f_0 < e^{-\lambda}$). Kondisi yang pertama lebih sering terjadi daripada yang kedua. Menurut Ridout *et al*. (1998), *excess zero* juga dapat mengakibatkan overdispersi. Besarnya proporsi data yang bernilai nol dapat berakibat pada ketepatan (presisi) dari pendugaan (Winkleman, 2008). Menurut Broek (1995), uji *excess zero* secara sederhana yaitu

$$\chi^2_{hit} = \frac{(n_0 - n\hat{p}_0)^2}{n\hat{p}_0(1-\hat{p}_0) - n\bar{x}\hat{p}_0{}^2} \qquad (2)$$

dengan H_0 adalah tidak terjadi *excess zero* dan H_1 adalah terjadi *excess zero* pada data, $\bar{x} = \hat{\lambda}$ adalah rataan dari data cacahan, n dan n_0 masing-masing adalah jumlah observasi dan jumlah nilai nol, n_0 kemudian dibandingkan dengan $n\hat{p}_0$ di mana $\hat{p}_0 = exp(\text{-}\hat{\lambda})$. Tolak H_0 jika statistik uji Persamaan (2) lebih besar dari pada nilai tabel sebaran χ^2 dengan derajat bebas 1.

Zero Inflated Poisson Mixed Model

Mixed model merupakan model dengan pengaruh tetap dan pengaruh acak (Stroup 2013). GLMM yang dibangun dari peubah respon y_i yang memiliki sebaran Poisson dapat ditulis sebagai:

$$\log(y_i) = \boldsymbol{x_i'\beta} + v_i \qquad (3)$$

dengan $\boldsymbol{\beta}$ merupakan pengaruh tetap, v_i merupakan pengaruh acak, $\boldsymbol{x_i}$ merupakan peubah-peubah penjelas yang menggambarkan pengaruh tetap.

Menurut Lambert dalam Cameron dan Trivedi (1998), jika data yang bernilai nol atau kosong dijumpai pada data cacahan dan berjumlah banyak (*excess zeros*) maka disarankan untuk menggunakan model regresi *Zero Inflated Poisson* (ZIP). Model ZIP merupakan campuran antara regresi Poisson dengan regresi logistik. Suatu individu pengamatan akan masuk dalam kelompok *A* yang selalu bernilai nol (*the Always-0 Group* atau *zero state*) dengan peluang p_i dan akan masuk ke dalam kelompok $\tilde{A}$ (*the Not Always-0 Group* atau *non-zero state*), di mana nilai nol dan positif pada data, keduanya dihasilkan oleh suatu fungsi sebaran untuk data cacahan dengan peluang $1 - p_i$. Fungsi massa peluang ZIP yaitu sebagai berikut:

$$\begin{aligned} f(y_i) &= p_i + (1-p_i)e^{-\mu_i}, \; y_i = 0 \\ f(y_i) &= \frac{(1-p_i)e^{-\mu_i}\mu_i{}^{y_i}}{y_i!}, \; y_i > 0 \end{aligned} \qquad (4)$$

ZIP *mixed model* didefinisikan sebagai berikut:

$$\log(\mu_i) = \boldsymbol{x_i'\beta} + v_i$$
$$logit\,(p_i) = log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{z_i'\delta} + u_i \quad (5)$$

dengan $\boldsymbol{x_i}$ dan $\boldsymbol{z_i}$ adalah peubah penjelas yang masing-masing mempengaruhi rataan Poisson pada kelompok $\tilde{A}$ dengan parameter $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$, dan memengaruhi peluang pada kelompok A dengan parameter $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_m)'$, v_i dan u_i merupakan efek acak area, sedangkan fungsi log di sini merupakan fungsi logaritma natural (ln). $\boldsymbol{x_i}$ dan $\boldsymbol{z_i}$ dalam penelitian ini menggunakan peubah yang sama. Rataan dan ragam dari model ZIP adalah

$$E(Y_i) = (1-p_i)\mu_i$$
$$V(Y_i) = (1-p_i)[\mu_i^2 + \mu_i] - (1-p_i)^2 \quad (6)$$

Pendugaan RMSE

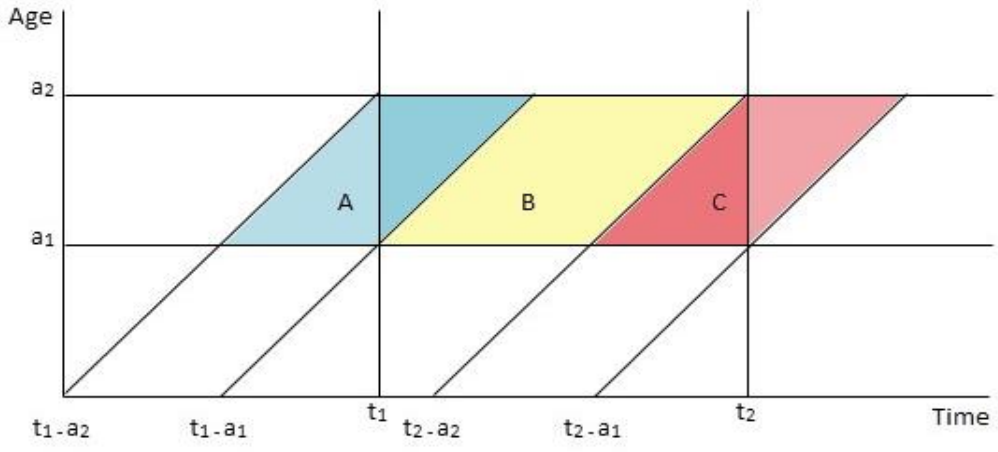
Mean Squared Error (MSE) model area kecil dapat dihitung dengan pendekatan *parametric bootstrap*. *Parametric bootsrap* merupakan salah satu alternatif perhitungan MSE yang mudah diaplikasikan pada pemodelan statistik yang rumit dan tidak terlalu bergantung kepada jumlah area kecil (Molina *et al*. 2009). Formula yang digunakan untuk RMSE yaitu sebagai berikut, *T* menyatakan jumlah *bootstrap* yang dibangkitkan.

$$RMSE_i = \left(\sqrt{T^{-1}\sum_{t=1}^{T}(\hat{\mu}_{it} - \mu_{it})^2}\right) \quad (7)$$

Konsep AKB

Berdasarkan laporan SDKI 2017, kematian bayi didefinisikan sebagai peluang kematian antara kelahiran dan sebelum mencapai ulang tahun pertama (0-1) tahun, sedangkan AKB merupakan kematian bayi tiap seribu kelahiran hidup. Pendugaan AKB menggunakan data pada tanggal kelahiran, status kelangsungan hidup, dan tanggal kematian anak atau usia pada saat anak meninggal. Penghitungannya menggunakan metode *synthetic cohort life table*. Metode ini memungkinkan untuk pendugaan trend AKB dari waktu ke waktu (USAID, 2018). Cara penghitungannya yaitu peluang kematian untuk tiap segmen umur ditabulasi. Segmen umur ini meliputi kematian umur 0-30 hari (0 bulan), 1-2 bulan, 3-5 bulan, 6-11 bulan, 12-23 bulan, 24-35 bulan, 36-47 bulan, dan 48-59 bulan. Peluang kematian untuk tiap segmen didefinisikan oleh periode waktu (*time*/t) dan interval umur (*age*/a), yang dapat digambarkan dalam diagram Lexis. Berdasarkan diagram Lexis ini, tiga kohort kelahiran anak disertakan, seperti yang ditunjukkan pada Gambar 1.

Pembilang untuk perhitungan peluang kematian adalah jumlah semua kematian kohort B, setengah dari kematian



Gambar 1. *Synthetic Cohort Life Table* dalam Diagram Lexis (USAID 2018)

kohort A, dan semua kematian kohort C jika periode berakhir pada tanggal wawancara, jika tidak maka setengah. Penyebut untuk perhitungan peluang kematian adalah semua anak yang masih hidup dari kohort B, setengah dari anak yang masih hidup dari kohort A, dan semua anak yang masih hidup dari kohort C jika periode berakhir pada tanggal wawancara, jika tidak maka setengah. Misal p_i menyatakan peluang kematian dengan i=1, …, 8 di mana i adalah segmen umur 0-30 hari (0 bulan) hingga 48-59 bulan maka peluang bertahan hidup adalah 1-p_i. Peluang kematian bayi adalah $1-(1-p_1)(1-p_2)(1-p_3)(1-p_4)$, sehingga AKB adalah peluang kematian bayi dikali 1000.

Berdasarkan USAID (2018), penghitungan ragam untuk survei demografi dan kesehatan menggunakan metode linierisasi Taylor untuk pendugaan yang berbentuk rataan dan proporsi. Penghitungan ragam untuk statistik yang lebih kompleks, seperti tingkat fertilitas dan tingkat mortalitas, menggunakan metode *resampling* Jackknife.

Sumber Data

Data yang digunakan dalam penelitian merupakan data sekunder yang berasal dari SDKI 2017 dan profil kesehatan provinsi 2016. Data SDKI digunakan untuk memeperoleh dugaan langsung AKB sedangkan profil kesehatan provinsi digunakan untuk memperoleh peubah penyerta level kabupaten/kota. Peubah penyerta yang digunakan yaitu proporsi balita dengan pneumonia, proporsi bayi dengan asi eksklusif, proporsi balita dengan gizi buruk per 1000 balita, proporsi penduduk dengan akses air minum layak, proporsi balita dengan imunisasi dasar lengkap.

Proses Analisis Data

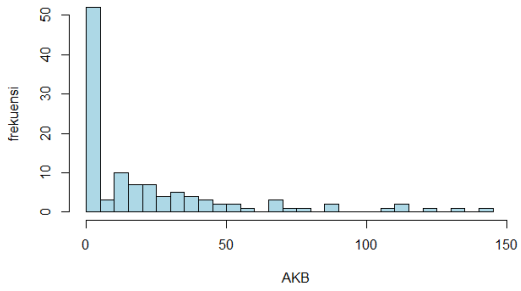
1. Melakukan eksplorasi data.
2. Menduga angka kematian bayi per kabupaten/kota menggunakan metode pendugaan langsung. Pendugaan ini menggunakan bantuan *package childhoodmortality* dalam program R.
3. Pemodelan angka kematian bayi dengan Poisson *mixed model*.
4. Melakukan pengecekan nilai dispersi dan kondisi *excess zero*.
5. Jika terjadi permasalahan dispersi dan *excess zero* maka dilakukan pemodelan alternatif. Pemodelan alternatif yang dilakukan yaitu *Zero Inflated Poisson mixed model*.
6. Menghitung RMSE dugaan model dengan teknik bootstrap.
7. Membandingkan RMSE untuk mendapatkan model terbaik. Model terbaik adalah model dengan RMSE terendah.
8. Menghitung nilai dugaan AKB kabupaten/kota berdasarkan model terbaik.

HASIL DAN PEMBAHASAN

SDKI 2017 menggunakan empat macam kuesioner, yaitu rumah tangga, wanita usia subur, pria kawin, dan remaja pria. Pengukuran angka kematian bayi diperoleh dari kuesioner wanita usia subur (WUS) umur 15-49 tahun dengan bayi yaitu anak berusia 0-12 bulan. Tidak semua kabupaten/kota memiliki ukuran sampel bayi yang cukup untuk pendugaan AKB level kabupaten/kota, bahkan terdapat kabupaten/kota dengan sampel bayi kurang dari lima puluh anak. Hal ini mengakibatkan pendugaan langsung AKB level kabupaten/kota akan menghasilkan standard error yang besar.

Pendugaan langsung AKB dilakukan terhadap 113 kabupaten/kota yang memiliki sampel. Penimbang yang digunakan yaitu penimbang sampel individu wanita yang disediakan pula dalam data SDKI. Gambar 2 merupakan histogram hasil pendugaan AKB berdasarkan SDKI untuk level kabupaten/kota. Hasil pendugaan AKB tersebut menunjukkan bahwa cukup banyak kabupaten/kota dengan AKB bernilai nol yaitu sebesar 45,13 persen. Meskipun hasil dugaan langsung AKB bernilai nol, belum tentu tidak terdapat kematian bayi di kabupaten/kota tersebut. Hal ini dapat disebabkan oleh ukuran sampel kecil sehingga WUS yang memiliki kematian bayi tidak terpilih sebagai sampel. Hasil pendugaan langsung AKB memiliki standard error yang besar. Oleh karena itu,

metode pendugaan tidak langsung dilakukan agar menghasilkan dugaan AKB dengan presisi yang lebih baik. Selain itu, pendugaan tidak langsung dilakukan melalui sebuah model. Model ini melibatkan peubah penyerta sebagai pengaruh tetap dan area kecil sebagai pengaruh acak, yang disebut dengan model campuran (mixed model).



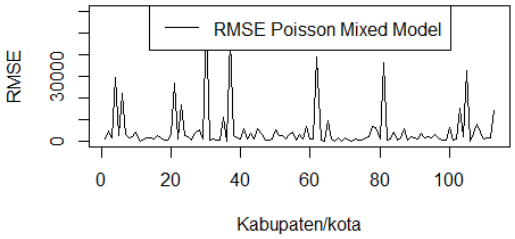
Gambar 2. Histogram angka kematian bayi per kabupaten/kota di Pulau Jawa

Peubah penyerta dalam pemodelan area kecil memiliki peran penting karena dapat mempengaruhi hasil pendugaan. Peubah penyerta yang digunakan adalah yang diduga berhubungan dengan angka kematian bayi berdasarkan penelitian-penelitian sebelumnya dan tersedia untuk seluruh kabupaten/kota di Pulau Jawa. Pertimbangan yang dijadikan dasar dalam pemilihan peubah penyerta yaitu peubah yang memiliki korelasi kuat dan signifikan. Peubah penyerta berasal dari data administratif dinas kesehatan, sehingga dianggap tidak memiliki error.

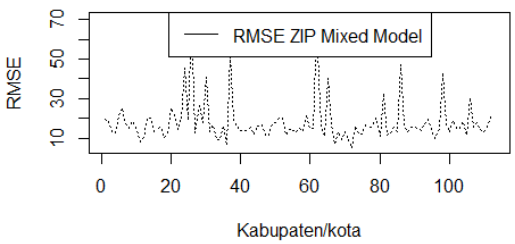
Pendugaan secara tidak langsung dilakukan dengan berbasis model. Angka kematian bayi merupakan data cacahan dengan peluang kejadian sangat kecil, sehingga dapat diasumsikan memiliki sebaran Poisson. Hampir 50 persen dugaan langsung AKB bernilai nol. Hal ini mengindikasikan adanya excess zero pada data. Excess zero merupakan salah satu penyebab terlanggarnya asumsi kesamaan rataan dan ragam pada pemodelan Poisson (equidispersi). Oleh karena itu dilakukan pendeteksian terhadap masalah dispersi dan excess zero.

Pengujian excess zero dilakukan dengan menggunakan fungsi zero.test pada package vcdExtra dalam program R. Hipotesis nol pada uji ini yaitu model poisson dapat menangani nilai nol dengan baik. Hasil pengujiannya menghasilkan p-value sebesar $2.22e^{-16}$. Hal ini berarti model Poisson tidak dapat menangani nilai nol dengan baik, dengan kata lain yaitu terjadi excess zero pada data. Ketika asumsi equidispersi dalam pemodelan Poisson terlanggar dan terjadi excess zero, pendugaan parameter menjadi bias dan uji statistik yang diturunkan dari model menjadi tidak benar. Oleh karena itu, pemodelan alternatif ZIP mixed model dicobakan dalam penelitian ini.

Semakin kecil RMSE maka semakin bagus suatu model dalam pendugaan. Gambar 3 menunjukkan bahwa Poisson mixed model menghasilkan RMSE yang sangat besar. RMSE Poisson mixed model mencapai nilai ribuan, sedangkan ZIP mixed model menghasilkan RMSE di bawah 70. Hal ini menunjukkan bahwa ZIP mixed model lebih baik dibandingkan Poisson mixed model. Selanjutnya, RMSE hasil pemodelan ZIP mixed model dibandingkan dengan RMSE dugaan langsung untuk membuktikan apakah dugaan model ini dapat memperbaiki dugaan langsung.



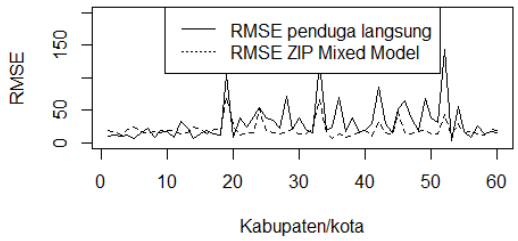
Gambar 3. RMSE Poisson *Mixed Model*



Gambar 4. RMSE ZIP *Mixed Model*

Perbandingan dengan dugaan langsung, RMSE ZIP *mixed model* lebih rendah dibandingkan dugaan langsung

(Gambar 5). Hal ini menunjukkan bahwa hasil dugaan ZIP *mixed model* dapat memperbaiki dugaan langsung dan dapat digunakan untuk menduga AKB level kabupaten/kota di Pulau Jawa. Dugaan AKB level kabupaten/kota di Pulau Jawa dapat dilihat di lampiran.



Gambar 5. Grafik RMSE dugaan langsung dan ZIP *Mixed Model*

KESIMPULAN

Berdasarkan hasil analisis pada angka kematian bayi diketahui terdapat masalah excess zero pada Poisson *mixed model*. Pemodelan alternatif yang dicoba yaitu ZIP *mixed model*. Pendugaan dengan ZIP *mixed model* menghasilkan RMSE lebih rendah dibandingkan dengan model Poisson *mixed model*. RMSE ZIP *mixed model* juga lebih rendah dibandingkan dengan penduga langsung. Oleh karena itu, ZIP mixed model dapat digunakan untuk menduga AKB level kabupaten/kota.

DAFTAR PUSTAKA

Anggreyani A. 2016. Kajian Pendugaan Area Kecil Untuk Menduga Jumlah Kematian Bayi Di Jawa Barat [tesis]. Bogor (ID): Institut Pertanian Bogor.

[BPS] Badan Pusat Statistik (ID). 2017. Laporan Survei Demografi dan Kesehatan 2017.

Broek J. 1995. A Score Test for Zero Inflation In a Poisson Distribution. *Biometrics*. 51: 738-743. doi: 10.2307/2532959.

Cameron A, Trivedi P. 1998. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Dean CB, Lundy ER. 2016. Overdispersion. Wiley StatsRef: Statistics Reference Online. doi: 10.1002/9781118445112.stat06788.pub2

Faraway JJ. 2016. Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models (2nd ed.). CRC Pr.

Kurnia A. 2009. Prediksi terbaik empirik untuk model transformasi logaritma di dalam pendugaan area kecil dengan penerapan pada data susenas [disertasi]. Bogor (ID): Institut Pertanian Bogor.

Molina I, Salvati N, Pratesi M. 2009. Bootstrap for Estimating the MSE of The Spatial EBLUP. *Comput Stat.* 24:441–458.

Murray CJL, Salomon JA, Mathers C. 2000. A Critical Examination of Summary measures of Population Health. *Bull World Health Organ*, 2000;78:981–94.

Rao JNK, Molina I. 2015. *Small Area Estimation 2nd ed*. New Jersey: John Wiley & Son.

Ridout M, Demetrio CGB, Hinde J. 1998. Models for Count Data With Many Zeros. *International Biometric Conference*, Cape Town, Desember 1998.

Stroup WW. 2013. *Generalized Linear Mixed Models*, New York: Chapman & Hall/CRC.

Winkelmann R. 2008. *Econometric Analysis of Count Data 5th edition*. Berlin: Springer.

[USAID] United States Agency for International Development (US). 2018. Guide to DHS statistics.

Winkelmann R. 2008. *Econometric Analysis of Count Data 5th edition*, Berlin: Springer.

Yudistira. 2018. Kajian metode pendugaan area kecil dalam pendugaan indikator bidang kebudayaan tingkat kabupaten/kota [tesis]. Bogor (ID): Institut Pertanian Bogor.

Lampiran Dugaan AKB Level Kabupaten/Kota di Pulau Jawa

kode kabupaten /kota	AKB dugaan langsung	AKB ZIP mixed model
3171	14	12
3172	18	17
3173	0	17
3174	15	15
3175	12	11
3201	18	18
3202	28	17
3203	46	26
3204	13	9
3205	31	21
3206	0	11
3207	0	13
3208	13	10
3209	9	6
3210	50	23
3211	0	18
3212	22	12
3213	0	12
3214	0	16
3215	7	8
3216	13	8
3217	29	14
3271	13	7
3272	125	64
3273	24	18
3274	106	54
3275	0	14
3276	9	10
3277	0	21
3278	0	39
3301	66	24
3302	43	18
3303	0	11
3304	0	12
3305	40	26
3306	0	8
3307	73	69
3308	68	34
3309	34	14
3310	0	18
3311	0	15
3312	22	9

kode kabupaten/kota	AKB dugaan langsung	AKB ZIP mixed model
3328	34	13
3329	15	10
3371	0	18
3372	0	19
3373	133	53
3374	18	10
3375	0	13
3376	111	61
3401	0	19
3402	23	5
3403	58	25
3404	17	6
3471	39	23
3501	0	12
3502	0	6
3503	0	20
3504	0	15
3505	0	15
3506	0	20
3507	26	25
3508	0	17
3509	20	18
3510	45	19
3511	145	51
3512	0	15
3513	0	15
3514	29	19
3515	16	8
3516	87	71
3517	0	17
3518	0	17
3519	55	28
3520	0	18
3521	0	19
3522	35	18
3523	19	13
3524	70	34
3525	38	19
3526	0	12
3527	22	10
3528	112	61
3529	0	20

kode kabupaten/ kota	AKB dugaan langsung	AKB ZIP mixed model
3313	0	17
3314	0	16
3315	0	17
3316	0	18
3317	0	13
3318	0	15
3319	51	32
3320	0	20
3321	21	15
3322	0	24
3323	0	14
3324	39	23
3325	0	14
3326	0	17
3327	0	18

kode kabupaten/kota	AKB dugaan langsung	AKB ZIP mixed model
3571	0	16
3572	0	25
3573	0	19
3575	0	17
3576	0	21
3578	4	6
3601	87	46
3602	42	25
3603	13	11
3604	33	16
3671	21	20
3672	0	18
3673	78	36
3674	15	16

PEMODELAN KERUGIAN BENCANA BANJIR AKIBAT CURAH HUJAN EKSTREM MENGGUNAKAN EVT DAN *COPULA*

Ian Surya Prayoga[1], Atina Ahdika [2]

[1,2] Program Studi Statistika, Fakultas MIPA, Universitas Islam Indonesia
e-mail: [1]iansurya17@gmail.com, [2]atina.a@uii.ac.id

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 $\hat{\theta}$ 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

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.

Keywords: *Extreme rainfall, floods, Extreme Value Theory, Copula*

PENDAHULUAN

Pada saat ini pola iklim global maupun regional telah mengalami banyak perubahan. Salah satu yang menyebabkan perbuhan pola iklim ini yaitu efek rumah kaca. Efek rumah kaca menyebabkan gas karbondioksida menumpuk pada lapisan ozon sehingga menyebabkan panas bumi yang berasal dari cahaya matahari tidak dapat keluar dari permukaan bumi. Akibatnya, suhu rata – rata permukaan bumi pada berbagai wilayah mengalami kenaikan. Kenaikan suhu rerata harian ini dapat mengakibatkan perubahan pada pola curah hujan yang menjadi tinggi bahkan menjadi ekstrem. Berdasarkan Peraturan Kepala Badan Meteorologi Klimatologi, dan Geofisika Nomor: Kep. 009 Tahun 2010, curah hujan ekstrem berupa hujan lebat dan hujan es. Hujan lebat adalah curah hujan dengan intensitas 20 mm/jam, sedangkan hujan es adalah curah hujan yang berbentuk butiran es dengan garis tengah paling rendah 5 mm. Curah hujan yang ekstrem ini dapat menjadikan berbagai bencana alam salah satunya bencana banjir. Contohnya ketika curah hujan ekstrem terjadi pada daerah pertanian akan menyebabkan petani mengalami penurunan produktifitas panen atau bahkan mengalami gagal panen . Sedangkan jika curah hujan ekstrem terjadi pada daerah padat penduduk yang wilayahnya kurang memiliki daerah resapan air, akan menyebabkan banjir yang kemudian bisa menyebabkan kerusakan bangunan seperti rumah warga atau bangunan pertokoan. Oleh karena itu, perlu dilakukan pemodelan untuk digunakan meramalkan kejadian curah hujan ekstrem agar bisa meminimalisirkan dampak yang disebabkan dari curah hujan ekstrem tersebut. Hasil peramalan ini juga bisa digunakan untuk mengoptimalkan bantuan – bantuan yang akan diberikan kepada pihak terdampak jika terjadi bencana alam.

Dalam ilmu statistika, salah satu metode yang digunakan untuk mengidentifikasi kejadian ekstreme yaitu dengan *Extreme Value Theory* (EVT). Kejadian yang bersifat ekstrem pada data *heavytail* dapat diramalkan menggunakan EVT. EVT juga dapat menjelaskan kerugian kejadian ekstrem yang kurang cocok dimodelkan dengan pendekatan biasa. *Extreme Value Theory* mengkaji perilaku stokastik variabel random secara minimum maupun maksimum (Kotz dan Nadarajah, 2000). Metode *extreme value theory* bertujuan untuk mengestimasi peluang suatu kejadian ekstrem dengan melihat ekor dari suatu distribusi berdasarkan nilai-nilai ekstrem yang diperoleh. Metode yang digunakan dalam mengidentifikasi suatu nilai ekstrem menggunakan *extreme value theory* yaitu metode *Block Maxima* (BM) dan metode *Peaks Over Threshold* (POT). Metode *Block Maxima* (BM) yaitu mengambil nilai maksimum dalam satu periode yang disebut sebagai blok dan metode *Peaks Over Threshold* (POT), yaitu mengambil nilai yang melewati suatu nilai threshold (McNeill, 1999).

Pada penelitian ini penulis akan melakukan pemodelan kerugian bencana banjir akibat curah hujan ekstrem di Provinsi Jawa Barat, Jawa Tengah dan Jawa Timur menggunakan *Extreme Value Theory* (EVT) dan pendekatan copula. EVT dalam penelitian ini digunakan untuk memodelkan curah hujan ekstrem. Curah hujan memiliki ekor distribusi yang gemuk (*heavytail*) sehingga perlu menggunakan EVT dalam pemodelannya. Sedangkan copula digunakan untuk menggambarkan ketergantungan antar variabel curah hujan dan kerugian bencana banjir. Dalam penelitian ini kerugian akibat banjir yang dimaksud yaitu rumah rusak. Tiga Provinsi tersebut merupakan salah satu daerah yang memiliki tingkat kasus bencana banjir yang tinggi, sehingga apabila terjadi hujan ekstrem yang berkesinambungan tentu saja berpengaruh terhadap kerusakan atau kerugian.

Berdasarkan rumusan masalah diatas, penelitian ini memiliki tujuan sebagai berikut:

a. Mendapat estimasi parameter pada pemodelan *Spatial Extreme Value* dengan pendekatan Copula spasial untuk curah hujan antara Jawa Barat, Jawa Tengah, dan Jawa Timur.

b. Mengetahui dependensi spasial curah hujan antara Provinsi Jawa Barat, Jawa Tengah dan Jawa Timur.
c. Mendapat model kerugian akibat bencana banjir di Provinsi Jawa Barat, Jawa Tengah dan Jawa Timur berdasarkan pemodelan *Spatial Extreme Value* dengan pendekatan Copula.

METODE

Curah Hujan

Menurut BMKG (BMKG, 2020), pengertian curah hujan adalah banyaknya curah hujan yang jatuh ke tanah dalam kurun waktu tertentu, dalam satuan milimeter (mm) di atas horizontal. Curah hujan juga dapat diartikan sebagai ketinggian air hujan yang tertampung dalam alat pengukur hujan, yang mempunyai permukaan datar, tidak menyerap, tidak menyerap, dan tidak mengalir. Nilai curah hujan 1 (satu) milimeter artinya satu milimeter air hujan atau satu liter air hujan ditampung di area seluas satu meter persegi di daerah datar. (BMKG, 2020).

Banjir

Banjir dapat diartikan sebagai tempat yang tergenang karena luapan yang meluap melebihi kapasitas pengolahan air suatu daerah atau tempat tertentu, dan menimbulkan banyak kerugian, seperti kerugian material, sosial dan ekonomi. Banjir juga dapat dijelaskan sebagai kejadian: lahan kering normal terendam karena curah hujan yang tinggi, dan biasanya area di wilayah tersebut rendah hingga cekung. Pada saat yang sama, banjir juga dapat disebabkan oleh limpasan limpasan, yang limpasannya melebihi kapasitas sistem drainase atau sistem aliran sungai. Selain itu, rendahnya kemampuan air untuk masuk ke dalam tanah juga dapat menyebabkan banjir yang menyebabkan tanah tidak dapat lagi menyerap air.

EVT (*Extreme Value Theory*)

EVT (*Extreme Value Theory*) adalah salah satu ilmu dalam statistika yang digunakan untuk melihat pola atau perilaku ekor (*tail*) dari suatu distribusi data ekstrem (Amalia, 2017). EVT digunakan untuk memodelkan peristiwa - peristiwa yang bersifat ekstrem, dimana peristiwa tersebut jarang terjadi dan berlangsung dalam waktu singkat akan tetapi peristiwa tersebut memberikan dampak kerugian yang cukup besar. EVT biasanya diterapkan pada kejadian yang besar dalam peristiwa alam seperti curah hujan, banjir, dan polusi udara.

Terdapat dua pendekatan yang digunakan untuk melakukan analisis menggunakan EVT yaitu dengan metode BM (*Block Maxima*) dan metode POT (*Peaks Over Threshold*). Metode BM (*Block Maxima*) yaitu pendekaran yang digunakan untuk mengidentifikasi nilai ekstrem dengan mengambil data atau nilai maksimum dari suatu observasi dalam satu periode tertentu yang disebut sebagai blok. Pendekatan ini hanya menghasilkan satu nilai ekstrem pada setiap blok. Sedangkan metode *Peaks Over Threshold* (POT), yaitu pendekatan yang mengidentifikasi nilai ekstrem dengan mengambil nilai yang melewati suatu nilai batas (*threshold*) tertentu (Rinaldi, 2016).

Distribusi *Generalized Extreme Value*

Metode *block maxima* mengaplikasikan teorema Fisher dan Tippet (1928) dalam Gilli dan Kellezi (2006), dimana dalam teorema tersebut menyatakan bahwa data sampel nilai ekstrem yang diambil dengan metode BM akan berdistribusi Gumble, Frechet atau Weibull. Kombinasi dari ketiga distribusi ini masuk ke dalam satu keluarga disebut sebagai distribusi *Generalized Extreme Value* (GEV). Menurut Mallor, Nualart, dan Omey (2009) *Generalized Extreme Value* (GEV) memiliki *cumulative distribution function* (CDF) seperti persamaan (1) berikut :

$$F(x;\mu;\sigma;\xi)=\begin{cases}\exp\left\{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right\} & ,-\infty<x<\infty,\xi\neq 0,-\infty<\mu<\infty,\sigma>0\\ \exp\left\{-\exp\left(\frac{x-\mu}{\sigma}\right)\right\} & ,-\infty<x<\infty,\xi=0,-\infty<\mu<\infty,\sigma>0\end{cases} \quad (1)$$

Probability distribution function (pdf) untuk distribusi GEV seperti persamaan (2).

$$f(x;\mu,\sigma,\xi)=\begin{cases}\frac{1}{\sigma}\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right)\right\}^{-\frac{1}{\xi}}\exp\left\{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right\} & ,\xi\neq 0,1+\xi\left(\frac{x-\mu}{\sigma}\right)>0\\ \frac{1}{\sigma}\exp\left\{\frac{x-\mu}{\sigma}\right\}\exp\left\{-\exp\left(\frac{x-\mu}{\sigma}\right)\right\} & ,\xi=0\end{cases}$$

(2)

Dimana x adalah nilai ekstrem yang diperoleh dari *block maxima* dengan $-\infty < x < \infty$, μ adalah parameter lokasi (location) dengan $-\infty < \mu < \infty$, σ adalah parameter skala (scale) dengan $\sigma > 0$ dan ξ adalah parameter bentuk (shape) dengan $-\infty < \xi < \infty$.

Uji Kesesuaian Distribusi

Tes Anderson Darling adalah tes yang digunakan untuk menentukan apakah data mengikuti distribusi tertentu. Tes Anderson Darling dengan sebuah program dapat digunakan untuk menguji kesesuaian distribusi GEV dengan data yang ekstrem (Engmann & Cousineau, 2011).

Berikut merupakan hipotesis pada Uji Anderson Darling

1. Uji Hipotesis :
   H_0: $F(x) = F^*(x)$ (Data mengikuti distribusi teoritis $F^*(x)$)
   H_1: $F\ x \neq F^*(x)$ (Data mengikuti distribusi teoritis $F^*(x)$)
2. Statistik Uji :
   $$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left(\ln\left(F^*(x_i)\right)+\ln\left(1-\left(F^*(x_{n+1-i})\right)\right)\right)$$
   (3)
   Keterangan:
   $F(x)$: fungsi distribusi kumulatif data sampel
   $F^*(x)$: fungsi distribusi kumulatif
   n : ukuran sampel
3. Daerah Kritis
   Tolak H_0 jika *p-value* < α atau jika nilai $AD_{hitung} > AD_{tabel}$
4. Kesimpulam :
   Kesimpulan didapatkan dengan cara membandingkan nilai AD_{hitung} dengan nilai AD_{tabel} atau dengan membandingkan nilai *p-value* dengan tingkat signifikasi (α).

Max Stable Process

Max Stable Processes (MSP) merupakan perluasan dari distribusi *multivariate extreme value* ke dimensi tak hingga (*infinite dimension*). Suatu fungsi distribusi G dikatakan max stable jika dan hanya jika G berdistribusi GEV (Ramadhani I. R., 2015). Dalam metode MSP, sampel diambil dari nilai maksimum (*Maxima*) pada setiap lokasi (proses spasial) (Cooley, Nyckah, & Naveau, 2007). Dalam konsep spasial ekstrem terdapat dua metode pendekatan yaitu *max-stable* dan *copula* (Davison, Padoan, & Ribatet, 2012).

Copula

Menurut teorema sklar, copula merupakan suatu fungsi yang menghubungkan fungsi distribusi multivariat dengan distribusi marginalnya (Nelsen & Flores, 2005). Copula terbagi menjadi dua macam families, yaitu elliptical copula dan archimedian copula. Untuk kasus *Spatial Extreme Model* copula yang dapat digunakan adalah elliptical copula kerena keluarga copula elliptical ini mampu menggambarkan kekuatan ketergantungan antara pasangan variabel spasial (AghaKouchak, Easterling, Hsu, Schubert, & Sorooshian, 2013). Copula yang termasuk dalam elliptical copula adalah *Gaussian copula* dan *Student's t-copula*. Sedangkan Copula yang termasuk dalam archimedian yaitu Copula Joe, Copula Frank, Copula Gumble dan Copula Clayton.

Selain copula dengan satu parameter tersebut, terdapat juga copula dengan dua parameter. Copula dengan dua parameter, merupakan gabungan dua copula satu parameter. Strukturnya yang lebih fleksibel memungkinkan koefisien ketergantungan ekor atas dan bawah berbeda (Ramadhani I. , 2019). Copula dengan dua parameter yaitu Copula BB1, BB6, BB7 dan BB8. Selain copula yang disebutkan di atas, terdapat juga copula yang dirotasi sebesar 90^o, 180^o dan 270^o dari copula Clayton, Gumbel, Joe, BB1, BB6, BB7 dan BB8. Ketika memutarnya 180^o, hal tersebut berarti diperoleh copula yang sesuai dengan copula

aslinya, sedangkan rotasi 90^o dan 270^o memungkinkan untuk pemodelan ketergantungan negatif yang tidak mungkin dilakukan dengan copula standar yang tidak dirotasi.

Kelas copula yang memungkinkan untuk berbagai macam struktur ketergantungan diberikan oleh keluarga Archimedean (Sch¨olzel & Friederichs, 2008). Copula ini dapat dibangun dengan generator ϕ sebagai berikut:

$$C_X(u,v) = \phi^{-1}(\phi(u)+\phi(v)) \quad ,0 \le u,1 \le v \qquad (4)$$

Pemilihan Model Terbaik

Salah satu cara menentukan model terbaik yaitu menggunakan AIC (Akaike's Information Criterion). Saat memilih model terbaik, pertimbangkan jumlah parameter dalam model. Semakin kecil nilai AIC, semakin baik dan layak untuk digunakan model tersebut. Menurut Ligas dan Banasick (2012) Akaike Information Criterion (AIC) didefinisikan dengan persamaan (5) sebagai berikut:

$$AIC = -2\ell_p(\hat{\beta}) + 2q \qquad (5)$$

Dimana $\ell_p(\hat{\beta})$ adalah fungsi *ln pairwise likelihood*. Maximum Pairwise Likelihood Estimation (MPLE) adalah metode estimasi parameter yang menggunakan fungsi densitas pairwise/berpasangan dari dua variabel. Prinsip dasar dalam MPLE adalah membuat turunan pertama terhadap masing-masing parameter lalu menyamakan dengan nol. Fungsi *ln pairwise likelihood* didefinisikan melalui persamaan (6) berikut ini:

$\ell_p(\hat{\beta})$ = fungsi *ln pairwise likelihood*

$$\ell_p(\hat{\beta}) = \sum_{i=1}^{k}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} \ln\left(f\left(u_{ji}, u_{ki}; \hat{\beta}\right)\right) \qquad (6)$$

Nilai dari *Akaike Information Criterion* (AIC) digunakan untuk menentukan model *trend surface* terbaik. Persamaan model *trend surface* sebagai berikut:

$$\begin{aligned}\hat{\mu}(j) &= \hat{\beta}_{\mu,0} + \hat{\beta}_{\mu,1} longitude(j) + \hat{\beta}_{\mu,2} latitude(j)\\ \hat{\sigma}(j) &= \hat{\beta}_{\sigma,0} + \hat{\beta}_{\sigma,1} longitude(j) + \hat{\beta}_{\sigma,2} latitude(j)\\ \hat{\xi}(j) &= \hat{\beta}_{\xi,0}\end{aligned} \qquad (7)$$

Model trend surface parameter $\hat{\mu}(p), \hat{\sigma}(p), \hat{\xi}(p)$ adalah semua kombinasi model dengan komponen spasial garis lintang (*latitude*) dan bujur (*latitude*). Mulai dari model parameter hanya dengan satu variabel penjelas sampai model parameter dengan model kuadratik. Keseluruhan kombinasi model berlaku untuk semua parameter $\hat{\mu}(p)$ dan $\hat{\sigma}(p)$, sedangkan untuk parameter $\hat{\xi}(p)$ diasumsikan konstan (Hakim, 2016).

Distribusi Poisson

Distribusi Poisson adalah distribusi diskrit, yang merupakan probabilitas keluaran sistematis dalam satuan standar tertentu x kali. Contoh aplikasi termasuk jumlah dering telepon per jam di kantor, jumlah goresan atau cacat pada permukaan produk, jumlah bakteri dalam kultur, dan penetapan nomor telepon yang salah. (Shrader, 1991). Fungsi kepadatan peluang untuk distribusi poisson dengan parameter λ adalah:

$$f(x:\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & x = 0,1,2,\ldots \\ 0, & x\ lainnya \end{cases} \qquad (7)$$

Distribusi Binomial Negatif

Distribusi Binomial Negatif adalah distribusi hasil percobaan bernoulli yang diulang sampai mendapatkan sukses ke-k. Fungsi Probabilitas distribusi Binomial Negatif adalah sebagai berikut:

$$\Pr(X = x) = \binom{x-1}{k-1}\theta^k(1-\theta)^{x-k}, x = k, k+1, k+2,\ldots \qquad (8)$$

Dengan Pr(X=x) adalah probabilitas terjadi sukses ke-k pada percobaan ke x dan θ merupakan probabilitas sukses dari setiap percobaan konstan.

MAPE (*Mean Absolute Percentage Error*)

MAPE digunakan untuk mengukur keakuratan nilai estimasi model, yang dinyatakan dalam bentuk persentase rata-rata kesalahan absolut, dan lebih banyak digunakan untuk membandingkan data

dengan rasio interval waktu yang berbeda (Robial, 2018). Skala pengkategorian nilai MAPE yang digunakan pada penelitian dan formula perhitungannya dijelaskan pada Tabel 1 dan persamaan (9) berikut:

$$MAPE = \left(\frac{1}{n}\right)\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad (9)$$

Keterangan:

A_t : Data aktual pada periode ke-t

F_t : Data peramalan pada periode ke-t

n : Jumlah periode peramalan

Seperti dijelaskan pada Tabel 1 suatu

Tabel 1. Ukuran Kesalahan

MAPE	Hasil Peramalan
<10%	Sangat Baik
10-20%	Baik
20-50%	Layak/Cukup
>50%	Buruk

model dikatakan memiliki kinerja sangat baik apabila memiliki nilai MAPE dibawah 10% dan memiliki kinerja baik apabila nilai MAPE berkisar antara 10% - 20% dan dikatakan layak apabila nilai MAPE berkisar antara 20% - 50% dan apabila lebih dari itu dikatakan berkinerja buruk (Razak, 2017). Akan tetapi MAPE memiliki kelemahan menjadi tak terbatas atau undefined jika ada nilai nol dalam seri.

MASE (*Mean Absolute Scaled Error*)

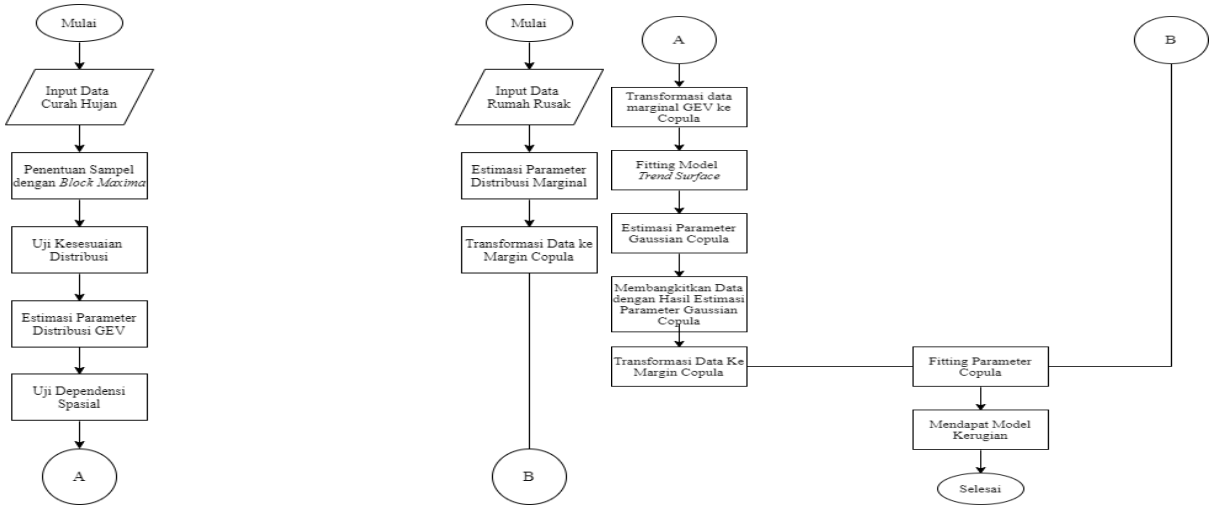
MASE dapat digunakan untuk membandingkan peramalan pada satu seri ataupun untuk membandingkan hasil perkiraan beberapa seri. Nilai tipikal untuk nilai satu-langkah MASE kurang dari satu, karena biasanya mungkin untuk mendapatkan perkiraan lebih akurat daripada metode naive.

$$MASE = \frac{mean|e_t|}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|} \quad (10)$$

Nilai MASE multistep sering lebih besar dari satu, karena menjadi lebih sulit untuk diperkirakan (Hyndman, 2006).

METODOLOGI PENELITIAN

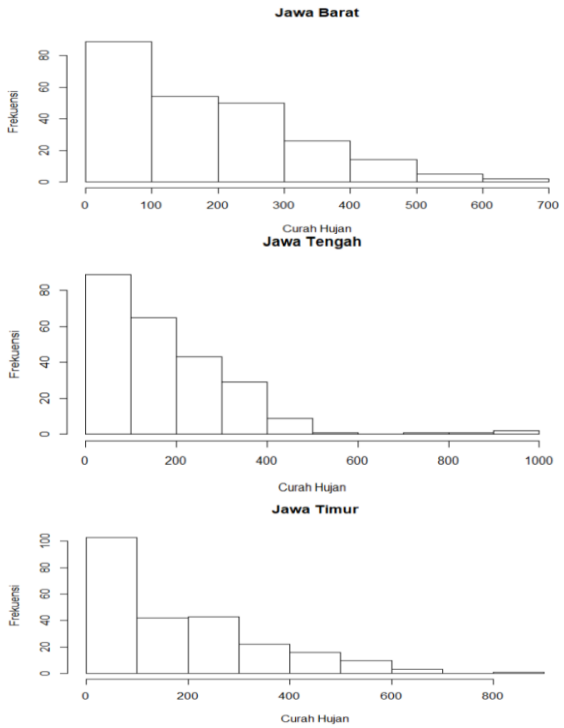
Penelitian ini menggunakan data sekunder yang diambil melalui 2 (dua) website. Data curah hujan diperoleh dari website dataonline.bmkg.go.id dan untuk data kerusakan rumah diperoleh dari website dibi.bnpb.go.id. Proses pengolahan data dilakukan dengan menggunakan software R 3.4.3. Sampel penelitian ini adalah data curah hujan bulanan dan data kerusakan rumah di Jawa Barat, Jawa Tengah dan Jawa Timur dari Maret 2000 hingga Desember 2019. Jumlah sampel yang diperoleh yaitu 242 data untuk masing-masing variabel. Setelah itu untuk menganalisis menggunakan extreme value theory, data sampel dikelompokkan menggunakan metode block maxima yaitu dengan membagi data ke dalam blok periode tertentu. Berdasarkan pengelompokan yang diacu dari BMKG, mengklasifikasikan pola hujan pada sebagian besar wilayah dipulau jawa merupakan pola hujan monsun, data dibagi ke dalam blok periode tiga bulanan, block



Gambar 1. Flowchart dari analisis penelitian ini

yang terbentuk yaitu Desember – Januari - Februari (DJF), Maret – April - Mei (MAM), Juni – Juli - Agustus (JJA) dan September - Oktober - November (SON). Selama periode sampel (2000-2019) terbentuk 79 blok. Dari satu blok diambil satu nilai ekstrem, nilai ekstrem yang diambil merupakan nilai maksimum dari masing-masing blok.

HASIL DAN PEMBAHASAN



Gambar 2. Distribusi Curah Hujan di Jawa Barat, Jawa Tengah, dan Jawa Timur

Melalui analisis visual menggunakan histogram terlihat adanya data ekstrem dan pola data *heavytail* pada ketiga Provinsi tersebut yang ditunjukan pada Gambar 2. Pada Gambar 2 terlihat bahwa pola dari distribusi data cenderung miring ke kiri , selain itu juga terlihat bahwa terdapat data yang memiliki frekuensi tinggi disekitar nilai nol, disisi lain juga terdapat data yang bernilai jauh lebih besar dari nilai nol akan tetapi memilik frekuensi yang sangat kecil, sehingga mengindikasikan adanya pola data *heavytail*. Karena terdapat indikasi nilai ekstrem dan data *heavytail* ini menunjukkan bahwa data curah hujan tidak berdistribusi normal dan mengakibatkan penelitian ini menggunakan metode *extreme value theory.*

Data Sampel dengan *Block Maxima*

Penentuan sampel dilakukan menggunakan metode *block maxima* yaitu dengan membagi data kedalam blok periode tertentu. Berdasarkan pengelompokan yang diacu dari BMKG, mengklasifikasikan pola hujan pada sebagian besar wilayah dipulau jawa merupakan pola hujan monsun, data dibagi ke dalam blok periode tiga bulanan, *block* yang terbentuk yaitu Desember – Januari - Februari (DJF), Maret – April - Mei (MAM), Juni – Juli - Agustus (JJA) dan September - Oktober - November (SON). Selama periode sampel (2000-2019) terbentuk 79 blok. Dari satu blok diambil satu nilai ekstrem, nilai ekstrem yang diambil merupakan nilai maksimum dari masing-masing blok.

Uji Kesesuaian Distribusi

Selanjutnya dilakukan pengujian *Anderson Darling* untuk mengetahui kesesuaian distribusi. Pengujian hipotesis sebagai berikut:

H_0: $F\,x = F^*\,x$ (Data mengikuti distribusi teoritis F^*x)

H_1: $F\,x \neq F^*\,x$ (Data mengikuti distribusi teoritis F^*x)

Statistik uji yang digunakan yaitu pada persamaan (3) dengan menggunakan tingkat signifikansi α =5% , tolak H0 jika AD_{hitung} lebih besar dari AD_{tabel}.

Tabel 2. Hasil Uji *Anderson Darling*

No	Provinsi	AD_{hitung}	P-value	Keputusan
1.	Jawa Barat	0.5653532	0.9344986	Gagal Tolak H_0
2.	Jawa Tengah	0.3106373	0.5876178	Gagal Tolak H_0
3.	Jawa Timur	0.9652988	0.9972561	Gagal Tolak H_0

Berdasarkan hasil pengujian menggunakan *Anderson Darling,* terlihat bahwa nilai p-value ketika provinsi tersebut lebih besar dari α sehingga menghasilkan keputusan gagal tolak H_0. Hal ini berarti bahwa data sampel ekstrem blok tiga bulanan sudah mengikuti distribusi GEV.

Estimasi Parameter GEV

Selanjutnya data sampel ekstrem curah hujan, digunakan untuk mengestimasi parameter distribusi GEV. Hasil estimasi parameter GEV disajikan pada Tabel 3.

Tabel 3. Hasil Estimasi Parameter Distribusi GEV

No	Provinsi	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$
1.	Jawa Barat	226.021 0656	141.5225 902	-0.2187897
2.	Jawa Tengah	179.299 86206	128.5039 3438	0.08802087
3.	Jawa Timur	181.837 45676	162.3772 1168	- 0.04369635

Selanjutnya untuk melakukan analisis *spatial* dengan pendekatan copula, data perlu dilakukan transformasi ke distribusi frechet karena distribusi frechet memiliki ekor yang paling *heavytail* dibandingkan distribusi GEV yang lain dan copula lebih tepat diterapkan untuk kasus *heavytail*. Setelah data ditransformasi ke frechet data perlu ditransformasi lagi ke copula.

Analisis Dependensi Spasial

Dalam penelitian ini sebanyak tiga lokasi atau wilayang yang diteliti. Hasil perhitungan koefisien ekstermal dapat dilihat pada tabel 4.

Tabel 4. Hasil Analisis Dependensi Spasial

No	Pasangan dua Provinsi	Jarak Euclid	Koefisien Ekstermal
1.	Jawa Barat - Jawa Tengah	2.823	1.424
2.	Jawa Tengah - Jawa Timur	2.403	1.027
3.	Jawa Barat - Jawa Timur	5.234	1.424

Nilai dari jarak euclid menggambarkan jarak antar lokasi dalam satuan derajat desimal. Satu derajat desimal sama dengan 111.319 km. Sehingga jarak euclid bernilai 2.83 sama dengan 2.83 x 111.319 yaitu 315.03277 km. Dari tiga pasang ini, nilai koefisien ekstermal berada pada rentang nilai 1.03- 1.42 sehingga dapat disimpulkan bahwa terdapat dependensi *spatial* antar lokasi. Dari hasil tersebut bisa disimpulkan bahwa ketiga Provinsi tersebut memiliki ketergantungan spasial.

Fitting Model *Trend Surface*

Pada penelitian ini, terdapat 9 kombinasi model *trend surface* dengan kombinasi variabel *longitude* dan *latitude*. Estimasi curah hujan ekstrem dilakukan menggunakan kombinasi model terbaik dari 9 kombinasi model yang ada. Hasil dari kombinasi model *trend surface* terdapat pada Tabel 5.

Tabel 5. Hasil Kombinasi Model *Trend Surface*

Kombi-nasi Ke	Kombinasi Model	AIC
1.	$\hat{\mu}(j)=\hat{\beta}_{\mu,0}+\hat{\beta}_{\mu,1}u(j)+\hat{\beta}_{\mu,2}v(j)$ $\hat{\sigma}(j)=\hat{\beta}_{\sigma,0}+\hat{\beta}_{\sigma,1}u(j)+\hat{\beta}_{\sigma,2}v(j)$ $\zeta(j)=\hat{\beta}_{\zeta,0}$	-90.1559
2.	$\hat{\mu}(j)=\hat{\beta}_{\mu,0}+\hat{\beta}_{\mu,1}u(j)+\hat{\beta}_{\mu,2}v(j)$ $\hat{\sigma}(j)=\hat{\beta}_{\sigma,0}+\hat{\beta}_{\sigma,1}u(j)$ $\zeta(j)=\hat{\beta}_{\zeta,0}$	-89.2062
3.	$\hat{\mu}(j)=\hat{\beta}_{\mu,0}+\hat{\beta}_{\mu,1}u(j)+\hat{\beta}_{\mu,2}v(j)$ $\hat{\sigma}(j)=\hat{\beta}_{\sigma,0}+\hat{\beta}_{\sigma,1}v(j)$ $\zeta(j)=\hat{\beta}_{\zeta,0}$	-92.2881
4.	$\hat{\mu}(j)=\hat{\beta}_{\mu,0}+\hat{\beta}_{\mu,1}u(j)$ $\hat{\sigma}(j)=\hat{\beta}_{\sigma,0}+\hat{\beta}_{\sigma,1}u(j)+\hat{\beta}_{\sigma,2}v(j)$ $\zeta(j)=\hat{\beta}_{\zeta,0}$	-97.6887
5.	$\hat{\mu}(j)=\hat{\beta}_{\mu,0}+\hat{\beta}_{\mu,1}v(j)$ $\hat{\sigma}(j)=\hat{\beta}_{\sigma,0}+\hat{\beta}_{\sigma,1}u(j)+\hat{\beta}_{\sigma,2}v(j)$ $\zeta(j)=\hat{\beta}_{\zeta,0}$	-92.7640
6.	$\hat{\mu}(j)=\hat{\beta}_{\mu,0}+\hat{\beta}_{\mu,1}u(j)$ $\hat{\sigma}(j)=\hat{\beta}_{\sigma,0}+\hat{\beta}_{\sigma,1}u(j)$ $\zeta(j)=\hat{\beta}_{\zeta,0}$	-92.5209
7.	$\hat{\mu}(j)=\hat{\beta}_{\mu,0}+\hat{\beta}_{\mu,1}u(j)$ $\hat{\sigma}(j)=\hat{\beta}_{\sigma,0}+\hat{\beta}_{\sigma,1}v(j)$ $\zeta(j)=\hat{\beta}_{\zeta,0}$	-93.2563
8.	$\hat{\mu}(j)=\hat{\beta}_{\mu,0}+\hat{\beta}_{\mu,1}v(j)$ $\hat{\sigma}(j)=\hat{\beta}_{\sigma,0}+\hat{\beta}_{\sigma,1}u(j)$ $\zeta(j)=\hat{\beta}_{\zeta,0}$	-93.0347
9.	$\hat{\mu}(j)=\hat{\beta}_{\mu,0}+\hat{\beta}_{\mu,1}v(j)$ $\hat{\sigma}(j)=\hat{\beta}_{\sigma,0}+\hat{\beta}_{\sigma,1}v(j)$ $\zeta(j)=\hat{\beta}_{\zeta,0}$	-91.4700

Berdasarkan tabel 5 diatas diketahui model yang terbaik adalah model ke-4 dengan AIC sebesar -97.6887. Dari model GEV terbaik dihitung nilai estimasi parameternya. Hasil estimasi patameter tersebut kemudian dimasukkan kedalam model sehingga diperoleh persamaan model *trend surface* terbaik sebagai berikut:

$\hat{\mu}(j) = 0.3145207 + 0.0006628u(j)$

$\hat{\sigma}(j) = 2.41652 + 0.11434v - 0.01206u(j)$

$\hat{\xi}(j) = -0.1449$

Estimasi Parameter *Gaussian Copula*

Nilai estimasi parameter copula untuk masing-masing lokasi curah hujan di Provinsi Jawa Barat, Jawa Tengah dan Jawa Timur disajikan dalam Tabel 6.

Setelah mendapatkan parameter copula gaussian dari hasil model *trend surface* terbaik, kemudian parameter itu digunakan untuk membangkitkan data random. Data random yang diperoleh seperti terdapat pada Tabel 7.

Tabel 6. Hasil Estimasi Parameter Gaussian Copula dengan Model *Trend Surface*

Provinsi	Latitude	Longitude	Lokasi ($\hat{\mu}$)	Skala ($\hat{\sigma}$)	Bentuk ($\hat{\zeta}$)
Jawa Barat	-6.8836	107.5973	0.386	0.332	-0.145
Jawa Tengah	-6.9486	110.4199	0.388	0.290	-0.145
Jawa Timur	-7.3846	112.7833	0.389	0.212	-0.145

Tabel 7. Data Random Menggunakan Gaussian Copula

No.	Jawa Barat	Jawa Tengah	Jawa Timur
1	0.606009757	0.55245563	0.31616379
2	2.274362693	1.51697878	1.58185776
3	1.056833761	2.31653719	3.13887328
4	295.1140986	119.768470	27.7540633
5	0.547848582	0.70725176	0.58033621
⋮	⋮	⋮	⋮
79.	0.524510522	1.49878442	1.683195015

Tabel 8. Prediksi Curah Hujan dengan Gaussian Copula

No	Jawa Barat		Jawa Tengah		Jawa Timur	
	Aktual	Prediksi	Aktual	Prediksi	Aktual	Prediksi
1	317.1	233.8	337.9	291.3	273.2	-9.9
2	526.4	219.0	326.7	944.1	589	92.4
3	563.8	379.1	454.6	211.9	571.2	169.7
4	457.7	415.1	593.6	426.5	640.7	296.7
5	304.8	82.8	381.5	128.8	556.2	117.9
6	416.7	200.4	270.7	205.2	318.4	388.1
7	299.9	467.1	809.5	456.1	886	434.6
8	499.8	-3.0	324.4	361.4	516.7	101.0
9	410.5	350.6	905	248.2	255.6	317.1
10	365.7	434.7	483.4	276.7	652.5	281.8
11	557.1	320.3	429.5	224.7	581.7	493.4
12	381.5	256.4	382.1	177.9	398.5	420.7
13	537	232.4	437.4	75.8	445.9	37.9
14	637	364.9	469	453.8	461.1	144.9
15	418.7	177.8	991.9	125.9	455.1	33.9
16	455	-2.0	261.8	382.4	479.8	214.7
17	311.5	271.9	324.3	204.4	589.6	165.0
18	389.3	642.0	404.5	117.8	400.7	106.1
19	483.2	467.9	708.6	205.9	528.5	128.9
20	322.9	187.1	372.8	232.2	487.8	69.8

Untuk membandingkan hasil prediksi curah hujan dengan data curah hujan yang asli, data random dari gaussian copula pada tabel 7 harus ditransformasikan kembali menjadi bentuk distribusi GEV terlebih dahulu. Hasil dari prediksi dapat dilihat pada Tabel 8.

Selanjutnya besaran error dari hasil prediksi curah hujan pada Tabel 8 dapat dilihat pada Tabel 9.

Tabel 9. Besar Error hasil Prediksi Curah Hujan dengan Gaussian Copula

Nomor	Provinsi	MAPE
1.	Jawa Barat	0.4484322
2.	Jawa Tengah	0.5410097
3.	Jawa Timur	0.6301803

Pada Tabel 9 terlihat bahwa nilai MAPE untuk Provinsi Jawa Barat sebesar 0.4484322, Provinsi Jawa Tengah sebesar 0.5410097 dan untuk Provinsi Jawa Timur yaitu 0.6301803. Jadi bisa disimpulkan bahwa model curah hujan untuk Provinsi Jawa Barat layak digunakan karena memiliki nilai MAPE 44.8 % dan model untuk Provinsi Jawa Tengah dan Jawa Timur dapat dikatakan buruk karena memiliki nilai MAPE 54% dan 63%.

Estimasi Parameter Data Rumah Rusak

Oleh karena data rumah rusak yang digunakaan memilik banyak nilai 0, maka distribusi yang bisa digunakan untuk melakukan estimasi parameter yaitu distribusi poisson dan negative binomial.

Dari Tabel 10 tersebut terlihat bahwa nilai AIC dari distribusi binomial negatif lebih kecil dari pada distribusi poisson, sehingga untuk data rumah rusak distribusi yang digunakan yaitu distribusi binomial negatif.

Fitting Copula

Dari Hasil *Fitting* keluarga copula untuk Provinsi Jawa Barat, Jawa Tengah dan Jawa Timur diperoleh hasil copula terbaik yang dapat dilihat Tabel 11.

Tabel 11. Besar Error hasil Prediksi Curah Hujan dengan Gaussian Copula

Provinsi	Copula	Parameter	AIC
Jawa Barat	Frank	1.49998	-1.50126
Jawa Tengah	Frank	-0.58169	1.23643
Jawa Timur	Frank	-0.86483	1.045856

Berdasarkan Tabel 11 diperoleh model ketergantungan antara curah hujan ekstrem dengan rumah rusak akibat banjir pada Provinsi Jawa Barat, Jawa Tengah dan Jawa Timur yaitu copula Frank. Bila variabel tersebut mengikuti copula Frank artinya memiliki hubungan yang erat ketika kedua variabel rendah atau kuat dilihat dari nilai parameternya.

Berdasarkan model yang terpilih, model masing – masing copula dapat ditulis seperti berikut ini:

Tabel 10. Hasil Estimasi Parameter Menggunakan Distribusi Poisson

Variabel	Distribusi	Parameter		AIC
		Parameter 1	Parameter 2	
Jawa Barat	Poisson	120.6329	-	34360.99
Jawa Tengah	Poisson	121.3544	-	35592.09
Jawa Timur	Poisson	257.8861	-	94807.33
Jawa Barat	Binomial Negatif	0.1060129	120.642237	605.7986
Jawa Tengah	Binomial Negatif	0.128063	38.2205640	638.5495
Jawa Timur	Binomial Negatif	0.06679072	258.07516612	526.7250

1. Model copula frank pada Jawa Barat yaitu:

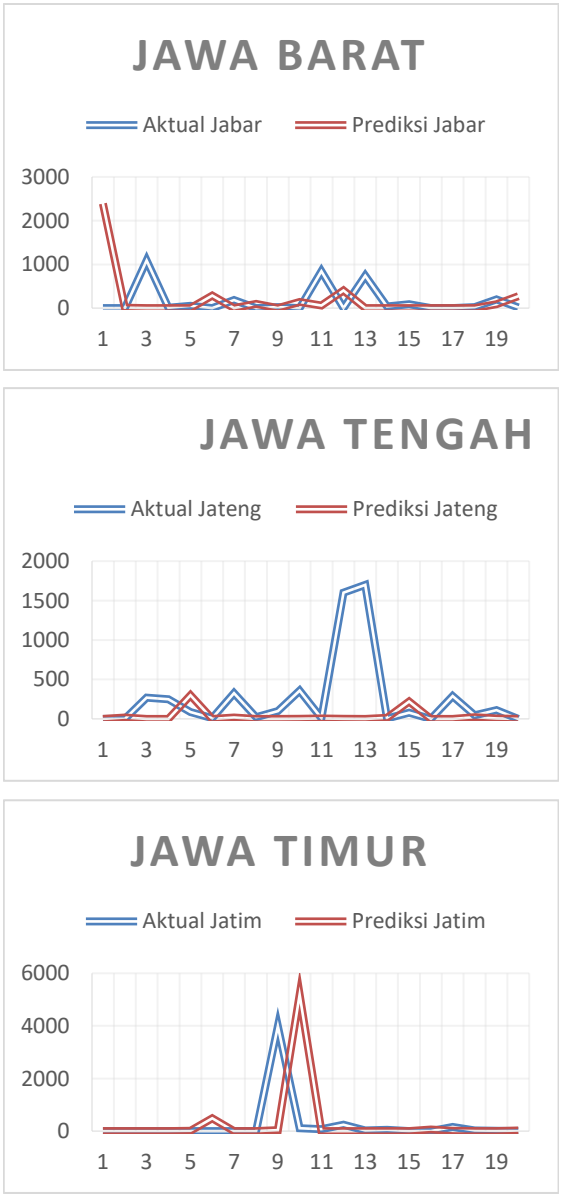
$$C_\theta(u,v) = -\frac{1}{\theta}\ln\left(1+\frac{\left(e^{\theta u}-1\right)\left(e^{\theta v}-1\right)}{e^{-\theta}-1}\right)$$
$$= -\frac{1}{1.49}\ln\left(1+\frac{\left(e^{1.49u}-1\right)\left(e^{1.49v}-1\right)}{e^{-1.49}-1}\right)$$

2. Model copula frank pada Jawa Tengah yaitu:

$$C_\theta(u,v) = -\frac{1}{\theta}\ln\left(1+\frac{\left(e^{\theta u}-1\right)\left(e^{\theta v}-1\right)}{e^{-\theta}-1}\right)$$
$$= \frac{1}{0.58}\ln\left(1+\frac{\left(e^{-0.58u}-1\right)\left(e^{-0.58v}-1\right)}{e^{0.58}-1}\right)$$

3. Model copula frank pada Jawa Timur yaitu:

$$C_\theta(u,v) = -\frac{1}{\theta}\ln\left(1+\frac{\left(e^{\theta u}-1\right)\left(e^{\theta v}-1\right)}{e^{-\theta}-1}\right)$$
$$= \frac{1}{0.86}\ln\left(1+\frac{\left(e^{-0.86u}-1\right)\left(e^{-0.86v}-1\right)}{e^{0.86}-1}\right)$$



Gambar 3. Prediksi Rumah Rusak dengan Copula Terbaik

Selanjutnya dari model yang telah diperoleh dilakukan random data menggunakan model copula terbaik yang kemudian hasilnya akan dibandingkan dengan data asli. Untuk membandingkan hasil prediksi rumah dengan data rumah rusak yang asli, data random dari copula terbaik harus ditransformasikan kembali menjadi bentuk distribusi binomial negatif terlebih dahulu. Hasil peramalan data rumah rusak dapat dilihat pada Gambar 3.

Selanjutnya dari hasil prediksi tersebut, dilakukan perhitungan error untuk mengetahui seberapa baik hasil prediksi.

Tabel 12. Besar Error Hasil Prediksi Rumah Rusak dengan Copula Terbaik

Nomor	Provinsi	MASE
1.	Jawa Barat	1.009497
2.	Jawa Tengah	0.84625
3.	Jawa Timur	1.112388

Dari Tabel 12 diperoleh nilai MASE untuk dari hasil prediksi pada Provinsi Jawa Barat yaitu 1.009497, untuk provinsi Jawa Tengah yaitu 0.84625 dan untuk Jawa Timur yaitu 1.112388.

KESIMPULAN

Beberapa simpulan sesuai dengan masalah yang dirumuskan, yakni antara lain:

1. Model curah hujan ekstreme pada Provinsi Jawa Barat, Jawa Tengah dan Jawa Timur menggunakan estimasi *Spatial Extreme Value* dengan pendekatan copula menghasilkan model *trend surface* sebagai berikut:

$$\hat{\mu}(j) = 0.3145207 + 0.0006628u(j)$$
$$\hat{\sigma}(j) = 2.41652 + 0.11434v - 0.01206u(j)$$
$$\hat{\xi}(j) = -0.1449$$

2. Nilai koefisien ekstermal pada ketiga pasang Provinsi berada pada rentang nilai 1.03- 1.42 sehingga dapat disimpulkan bahwa terdapat dependensi *spatial* antar Provinsi, artinya ketiga Provinsi tersebut memiliki ketergantungan spasial.
3. Model Copula untuk Provinsi Jawa Barat, Provinsi Jawa Tengah dan Jawa Timur mengikuti copula Frank. Nilai

parameter $\hat{\theta}$ copula Frank pada Provinsi Jawa Barat sebesar 1.4999840280, Provinsi Jawa Tengah sebesar -0.5816995330 dan Provinsi Jawa Timur sebesar -0.8648329345.

DAFTAR PUSTAKA

AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., & Sorooshian, S. (2013). *Extremes in a Changing Climate: Detection, Analysis and Uncertainly*. New York: Springer.

Amalia, L. F. (2017). Estimasi Parameter pada Pemodelan Spatial Extreme Value Dengan Pendekatan Copula.

BMKG. (2020). Diambil kembali dari Daftar Istilah Klimatologi: http://balai3.denpasar.bmkg.go.id/daftar-istilah-musim

Coles, S. (2001). Improving the Analysis of Extreme Wind Speed with Information-sharing Models. *de l'Institut Pierre Simon Laplace, no. 11, p. 12*, 284.

Cooley, D., Nyckah, D., & Naveau, P. (2007). A dependence measure for multivariate and spatial extremes: Properties and inference. *Journal of the American Statistical Association*, 824-840.

Davison, A., Padoan, S., & Ribatet, M. (2012). Statistical Modeling of Spatial Extremes. *Statistical Science*, 161-186.

Engmann, S., & Cousineau, D. (2011). *Jurnal of Applied Quantitative Methods*, Vol 6, No 3.

Hakim, A. R. (2016). *Pemodelan Spatial Extreme Value dengan Pendekatan Max-Stable Process.* Surabaya: Institute Teknologi Sepuluh Nopember.

Hyndman, R. J. (2006). Another Look At Forecast-Accuracy Metrics For Intermittent Demand. *FORESIGHT*, 43-46.

Nelsen, R. B., & Flores, M. Ú. (2005). The lattice-theoretic structure of sets of bivariate copulas and quasi-copulas. *Comptes Rendus Mathematique*, vol.341, no.9, hal.583-586.

Ramadhani, I. (2019). Identifikasi Struktur Dependensi dan Prediksi Indeks Harga Saham Gabungan Menggunakan Regresi Berbasis D-Vine Copula. Yogyakarta: Universitas Islam Indonesia.

Ramadhani, I. R. (2015). *TESIS - SS 142501*.

Razak, M. A. (2017). Peramalan Jumlah Produksi Ikan Dengan Menggunakan Backprogation Neural Network (Studi Kasus: UPTD Pelabuhan Perikanan Banjarmasin). Surabaya: Intitute Teknologi Sepuluh Nopember.

Rinaldi, A. (2016). Sebaran Generalized Extreme Value (GEV) dan Generalized Pareto (GP) untuk Pendugaan Curah Hujan Ekstrem di Wilayah DKI Jakarta. *Al-Jabar: Jurnal Pendidikan Matematika Vol. 7, No. 1*, Hal 75 - 84.

Sch¨olzel, C., & Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research – introduction to the copula approach. *Nonlinear Processes in Geophysics*.

PENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN JUMLAH KASUS COVID-19 DAN FASILITAS KESEHATAN

Meinisa Fadillah Rahmi[1], Paulus Satria Prasetyo[2], Ratih Nurhabibah[3], Rizky Perdana [4], [5]Wa Ode Zuhayeni Madjida[5]
[1,2,3,4]Politeknik Statistika STIS, [5]Badan Pusat Statistik
e-mail: [1]211709827@stis.ac.id

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

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*

PENDAHULUAN

Coronavirus Disease-2019 (COVID-19) pertama kali ditemukan pada Desember 2019 di kota Wuhan, Cina yang kemudian menyebar ke seluruh dunia sehingga ditetapkan sebagai pandemi global oleh *World Health Organization*. Penyebaran awal di Indonesia terjadi pada awal Maret 2020, dengan kasus pertama di Kota Depok, Jawa Barat. Virus corona menyebar melalui berbagai macam cara, diantaranya melalui interaksi antara penderita dengan orang lain, benda-benda yang disentuh oleh penderita, bahkan dapat menyebar melalui udara. Tercatat hingga 14 November 2020 jumlah kasus terkonfirmasi positif di Indonesia mencapai 463.007 jiwa dengan total jumlah pasien sembuh sebanyak 388.094 jiwa, sedangkan jumlah kematian sebanyak 15.148 jiwa.

Menurut Muhyiddin (2020), berbagai negara telah melakukan kebijakan *lockdown* untuk membatasi penyebaran virus. Akan tetapi, mengubah kebiasaan masyarakat bukan persoalan mudah, banyak negara mengalami kendala dalam penerapan kebijakan tersebut. Sehingga negara-negara tersebut memodifikasi kebijakan *lockdown* dengan menerapkannya secara penuh, sebagian, ataupun dalam skala lokal dan seminimal mungkin. Indonesia memodifikasi kebijakan *lockdown* menjadi Pembatasan Sosial Berskala Besar (PSBB) yang diberlakukan pada setiap provinsi atau kabupaten/kota berdasarkan tingkat keparahan yang dinilai oleh Kementerian Kesehatan. Penyebaran yang begitu cepat dan jumlah kasus terkonfirmasi yang terus mengalami kenaikan membuat pemerintah menerapkan berbagai kebijakan, mulai dari *physical distancing*, penggunaan masker, *work from home*, hingga pembelajaran jarak jauh guna memutus penyebaran mata rantai virus tersebut.

Ketersediaan infrastruktur dan layanan kesehatan yang terdapat pada masing-masing provinsi menentukan kesiapan daerah dalam penanganan pasien terkonfirmasi dan penyebaran berkelanjutan COVID-19. Berdasarkan data Ikatan Dokter Indonesia (IDI), jumlah dokter di Indonesia pada tahun 2020 sebanyak 183.605 dokter. Secara umum, jumlah tersebut mencukupi untuk seluruh masyarakat. Namun, permasalahannya terjadi pada persebaran yang tidak merata. Persebaran dokter bertumpuk di kota-kota besar dan Pulau Jawa, sehingga kebutuhan dokter di berbagai wilayah Indonesia masih belum terpenuhi. Begitupun dengan infrastruktur kesehatan dan tenaga medis lainnya. Ditambah lagi, hingga penghujung tahun 2020, belum terdapat kepastian mengenai vaksin untuk mencegah penyebaran COVID-19.

Noviyanto (2020) telah melakukan pengelompokkan jumlah kematian penderita COVID-19 di Benua Asia dengan menerapkan teknik *data mining*. Hasil dari penelitian ini menunjukkan jumlah kematian penderita COVID-19 dibagi kedalam 3 *cluster*. Terdapat 4 negara yang masuk kedalam *cluster* dengan tingkat kematian tinggi, yaitu Turki, Iran, India dan China. Negara yang masuk kedalam *cluster* dengan tingkat kematian sedang sebanyak 4 negara, yaitu Pakistan, Indonesia, Jepang, dan Piliphina. Terakhir, negara yang masuk kedalam *cluster* dengan tingkat kematian rendah adalah 41 negara lainnya.

Penelitian Rembulan dkk. (2020) mengenai kebijakan Pemerintah terkait COVID-19 di setiap Provinsi dengan menggunakan analisis *cluster* menghasilkan 3 *cluster* optimal. *Cluster* 1 adalah *cluster* dengan risiko tinggi karena memiliki variabel jumlah kasus aktif dan jumlah kasus kematian per satu juta penduduk yang tertinggi. *Cluster* 2 adalah *cluster* dengan risiko rendah karena memiliki variabel dengan jumlah kasus kesembuhan tertinggi dan jumlah kasus aktif terendah. *Cluster* 3 adalah *cluster* dengan risiko sedang karena memiliki variabel jumlah kesembuhan terendah dan jumlah kasus aktif sedang. Kebijakan pemerintah pada *cluster* 1 hendaknya memprioritaskan variabel jumlah kasus aktif dan jumlah kasus kematian per satu juta penduduk. Untuk *cluster* 2 harus memprioritaskan variabel jumlah kasus kematian, sedangkan untuk *cluster* 3 harus

memprioritaskan variabel jumlah kasus aktif.

Berdasarkan pemaparan di atas, sangat diperlukan penelitian mengenai pengelompokkan provinsi di Indonesia untuk mengidentifikasi tingkat penyebaran dan kapasitas penanganannya di masing-masing daerah. Oleh karena itu, penelitian ini bertujuan untuk mengelompokkan 34 provinsi berdasarkan jumlah kasus COVID-19 dan fasilitas kesehatan, menentukan metode yang terbaik, dan mengidentifikasi karakteristik masing-masing kelompok berdasarkan metode terbaik. Diharapkan dengan dibentuknya kelompok-kelompok tersebut dapat membantu pemerintah dalam menentukan kebijakan dan langkah yang harus diambil kedepannya guna menangani penyebaran COVID-19 di Indonesia.

METODOLOGI

1. Tinjauan Referensi

Data Mining

Data Mining adalah sebuah proses pencarian informasi secara otomatis dalam tempat penyimpanan data yang berukuran besar untuk menemukan pola tertentu. Teknik yang digunakan pada *data mining* tersebut dapat dikelompokkan ke dalam dua kategori, yaitu *descriptive mining* dan *predictive mining. Descriptive mining* adalah proses menemukan karakteristik tertentu dari basis data, meliputi *clustering, asosiation,* dan *sequential mining. Predictive mining* adalah proses menemukan pola dari data menggunakan variabel-variabel lain di masa depan, meliputi *classification.*

Clustering

Clustering merupakan pengelompokkan data untuk membentuk kelas baru dengan mengidentifikasi kelompok-kelompok yang memiliki kesamaan karakteristik tertentu. Objek-objek pada kelompok (*cluster*) yang terbentuk akan memiliki kemiripan satu sama lain dan berbeda dengan objek pada kelompok lain. Terdapat beberapa algoritma pada *clustering* diantaranya *K-Means, Fuzzy C-Means, DBSCAN,* dan *K-Medoids.* Setiap algoritma memiliki fungsi yang berbeda dan kelebihan serta kekurangannya masing-masing.

K-Means

K-Means Clustering merupakan saah satu algoritma dalam *clustering* yang paling sederhana. *K-means clustering* melakukan proses pemodelan menggunakan *unsupervised learning* dan mengelompokkan data menggunakan sistem partisi. Secara umum, terdapat dua jenis *clustering* yang digunakan dalam proses pengelompokkan data, yaitu *Hierarchical* dan *Non-Hierarchical. K-Means* merupakan salah satu metode *clustering Non-Hierarchical* atau *Partinional Clustering* yang secara umum dilakukan dengan algoritma sebagai berikut:

1. Menentukan jumlah *cluster* (k).
2. Menentukan titik pusat (*centroid*) *cluster* awal secara *random*.
3. Menghitung jarak *centroid* dengan setiap data. Ukuran jarak yang digunakan adalah *Euclidean distance* dengan rumus berikut.

$$D\ (i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \cdots + (X_{ki} - X_{kj})^2} \quad (1)$$

   Dimana D (i,j) = Jarak data ke-i ke pusat cluster j, X_{ki} = Data ke-i pada atribut data ke-k , X_{kj} = *Centroid* ke-j pada atribut ke-k
4. Data akan dimasukkan ke dalam *cluster* berdasarkan jarak *centroid* terdekat. Lalu hitung kembali *centroid cluster* yang baru. *Centroid cluster* merupakan rata-rata semua data dalam sebuah *cluster*, dapat dihitung menggunakan rumus berikut.

$$V\ (i,j) = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2)$$

   dimana: V (i,j) = *Centoid* rata-rata pada *cluster* i untuk atribut data ke-j

   N_i = Jumlah anggota *cluster* i
5. Ulangi penghitungan jarak *centroid* dengan setiap data (kembali pada langkah ke-3) jika masih terdapat data

yang berpindah *cluster* atau apabila perubahan nilai centroid berada di atas nilai *treshhold* yang dibentuk atau jika perubahan nilai pada *objective function* yang digunakan di atas nilai *threshold* yang ditentukan.

Kembali ke langkah 3, jika masih terdapat data yang berpindah *cluster* atau apabila perubahan nilai centroid berada di atas nilai *threshold* yang dibentuk atau jika perubahan nilai pada *objective function* yang digunakan di atas nilai *threshold* yang ditentukan.

K-Means with Outlier Detection

Menurut Chawla dan Gionis (2013), algoritma *k-means* sangat sensitif terhadap adanya *outlier* dalam data dikarenakan menggunakan rata-rata sebagai pusat *cluster*-nya sehingga seharusnya selain melakukan *clustering*, harus dilakukan deteksi
outlier juga dalam prosesnya. Hal ini dapat menghasilkan pengelompokkan yang lebih tepat. Metode ini menggunakan metode *minimum covariance determinant* (MCD) untuk menentukan *outlier*.

2. Metode Analisis

Penelitian ini bersifat kuantitatif yang mencakup 34 Provinsi di Indonesia dengan menggunakan data sekunder yang terdiri dari data kasus COVID-19 dan data fasilitas kesehatan. Data kasus COVID-19 diperoleh dari *website* resmi COVID-19 oleh Gugus Tugas dengan *update* kasus pada tangal 14 November 2020. Sedangkan data infrastruktur kesehatan diperoleh dari publikasi Rekapitulasi SDM Kesehatan yang Didayagunakan di Fasyankes pada Tahun 2019 oleh Kementerian Kesehatan.

Proses pengolahan data dilakukan dengan bantuan *software Microsoft Excel 2019* dan *RStudio.* Sebelumnya, data yang didapat dilakukan tahapan *preprocessing* untuk mempersiapkan variabel-variabel yang akan digunakan. Tahapan *preprocessing* yang dilakukan adalah sebagai berikut.

1. Melakukan *data cleaning* meliputi pengecekkan *missing data, outlier*, dan inkonsistensi. Data yang digunakan bebas dari *missing data* dan inkonsistensi, namun banyak terdapat *outlier*. Dalam kasus ini, penulis mempertahankan *outlier* dikarenakan masih ingin mempertahankan nilai yang sebenarnya pada tiap-tiap variabel dan diakomodasi dengan metode *k-means with outlier detection*.
2. Melakukan *data reduction*, yaitu mereduksi variabel yang tidak perlu dan melakukan agregat pada variabel-variabel yang menunjukkan karakteristik yang sama.
3. Melakukan *data transformation*, yaitu mengubah nilai variabel yang berupa jumlah infrastruktur menjadi rasio infrastruktur terhadap 1.000 penduduk yang ada di provinsi yang bersangkutan agar lebih menggambarkan kondisi yang sebenarnya dan lebih mudah dibandingkan, lalu melakukan normalisasi data menggunakan *z-score* dikarenakan variabel-variabel yang digunakan memiliki satuan yang berbeda-beda.

Berdasarkan hasil *preprocessing* yang telah dilakukan, didapatkan sebanyak 11 variabel yang siap dilakukan analisis, yaitu rasio jumlah puskesmas terhadap 1.000 penduduk (V1), rasio jumlah rumah sakit terhadap 1.000 penduduk (V2), rasio jumlah klinik terhadap 1.000 penduduk (V3), rasio jumlah posyandu terhadap 1.000 penduduk (V4), rasio jumlah dokter terhadap 1.000 penduduk (V5), rasio jumlah perawat terhadap 1.000 penduduk (V6), rasio jumlah bidan terhadap 1.000 penduduk (V7), jumlah kasus terkonfirmasi positif COVID-19 (V8), *recovery rate* (V9), *fatality rate* (V10), dan jumlah kasus aktif atau belum sembuh yang masih dalam perawatan atau isolasi mandiri (V11) menurut provinsi.

Tahapan analisis yang dilakukan selanjutnya adalah mengecek asumsi nonmultikolinearitas pada variabel-variabel yang digunakan dengan menghitung matriks korelasi antarvariabel. Kemudian, menentukan jumlah optimal *cluster* dengan menggunakan *by majority rule*, yaitu menggunakan *package NbClust* dimana terdapat 30 indeks yang melakukan *voting*

terhadap jumlah *cluster* yang optimal. Lalu, melakukan pengelompokkan objek dengan metode *k-means* dan *k-means with outlier detection* dengan ukuran ketidaksamaan berupa jarak Euclidean. Kemudian, menentukan metode terbaik dengan membandingkan nilai *R-squared* yang dihitung dari pembagian nilai varians *between* dengan varians total hasil *clustering*. Menurut Jhonson (2007), semakin besar nilai *R-squared*, maka semakin baik pengelompokkan yang dilakukan. Terakhir, melakukan identifikasi karakteristik masing-masing *cluster* dengan menghitung *centroid* berdasarkan metode terbaik.

HASIL DAN PEMBAHASAN

Setelah dilakukan tahapan *preprocessing data* dan didapatkan variabel-variabel yang siap digunakan untuk analisis, dilakukan tahap analisis *cluster*.

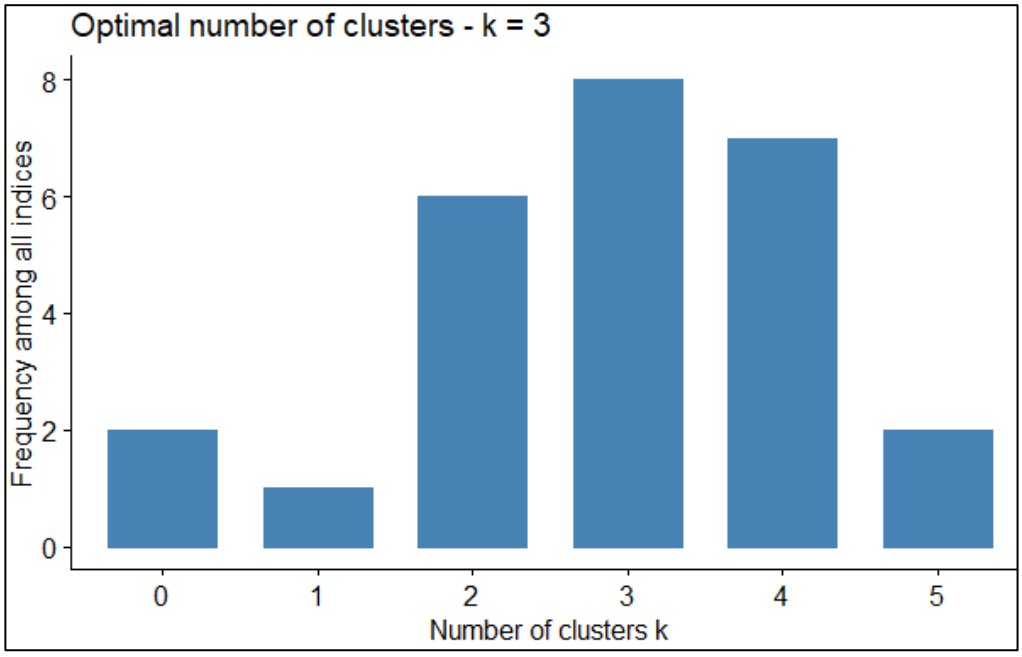
Pengujian Asumsi Nonmultikolinearitas

Analisis *cluster* harus memenuhi asumsi nonmultikolinearitas agar tidak ada variabel yang bersifat duplikasi atau merepresentasikan hal yang sama. Oleh karena itu, dilakukan pengujian dengan cara menghitung matriks korelasi antarvariabel. Didapatkan bahwa nilai korelasi mutlak antarvariabel yang digunakan memiliki rentang antara 0,00 hingga 0,70. Nilai korelasi terendah terdapat antara V2 dan V4. Sedangkan, nilai korelasi tertinggi terdapat antara variabel V8 dan V11. Dengan demikian, tidak terdapat nilai korelasi antarvariabel yang melebihi |0,8|. Artinya, variabel-variabel yang digunakan telah memenuhi asumsi nonmultikolinearitas dan dapat dilanjutkan ke tahap selanjutnya.

Penentuan Jumlah Cluster Optimal

Dalam pembentukan *cluster* ini, digunakan ukuran ketidaksamaan antarobjek berupa jarak Euclidean. Selanjutnya, dilakukan pemilihan jumlah *cluster* yang optimal untuk data yang digunakan menggunakan *by majority rule* dengan *package NbClust* di *software R*. Didapatkan hasil pada Gambae 1.

Berdasarkan hasil penentuan jumlah *cluster* yang optimal seperti tergambar pada Gambar 1, diperoleh bahwa jumlah *cluster* yang paling banyak direkomendasikan adalah 3 dengan jumlah *voting* sebanyak 8. Dengan kata lain, jumlah *cluter* yang optimal dalam kasus ini adalah 3. Oleh karena itu, akan dibentuk *cluster* yang berjumlah 3 untuk mengelompokkan provinsi-provinsi di Indonesia.



Gambar 1. Penentuan Jumlah *Cluster* Optimal

Pengelompokan Provinsi

Dalam penelitian ini, pengelompokkan provinsi dilakukan dengan menggunakan dua metode, yaitu metode *k-means* dan *k-means with outlier detection* untuk dibandingkan sehingga diperoleh metode yang terbaik untuk kasus ini.

Untuk metode *k-means*, diperoleh hasil pengelompokkan yang dilakukan dengan jumlah *cluster* 3, yaitu 6 provinsi di *cluster* 1, yaitu Bangka Belitung, Kepulauan Riau, Sulawesi Utara, DKI Jakarta, DI Yogyakarta, dan Bali. Sedangkan, *cluster* 2 ada 12 provinsi, yaitu Sumatera Utara, Riau, Sumatera Selatan, Lampung, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Selatan, dan Sulawesi Barat. Kemudian, 16 provinsi lainnya di *cluster* 3, yaitu Aceh, Sumatera Barat, Jambi, Bengkulu, Nusa Tenggara Timur, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua Barat, dan Papua.

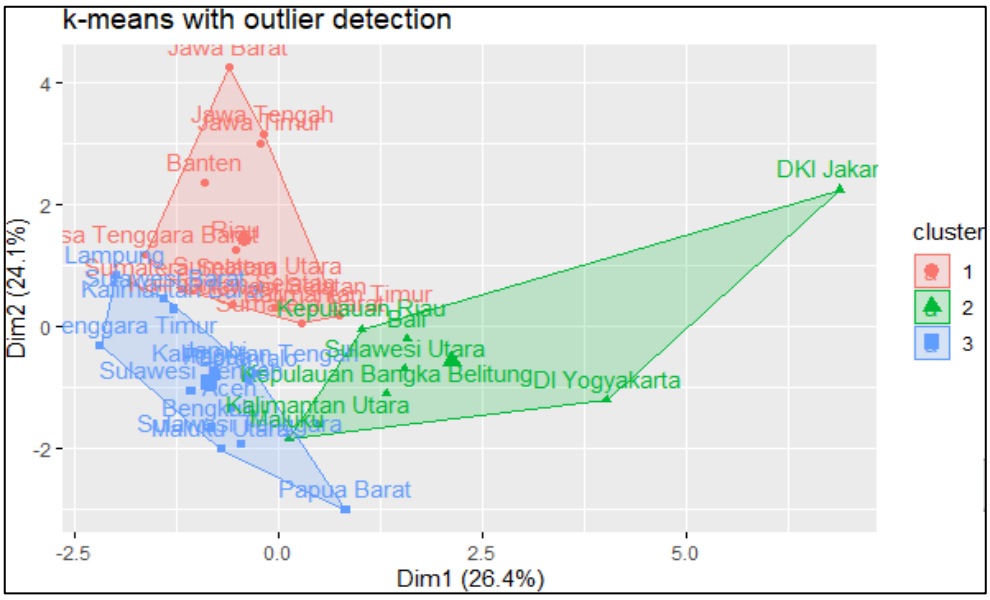
Hasil pengelompokkan dengan menggunakan metode *k-means with outlier detection* dengan jumlah *cluster* 3 dan *specified outlier* sebanyak 4 adalah terdapat 12 provinsi di *cluster* 1, yaitu Sumatera Utara, Sumatera Barat, Riau, Sumatera Selatan, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Barat, Kalimantan Selatan, Kalimantan Timur, dan Sulawesi Selatan. Sedangkan, 8 provinsi di *cluster* 2 adalah Bangka Belitung, Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Bali, Kalimantan Utara, Sulawesi Utara, dan Maluku. Kemudian, 14 provinsi lainnya masuk ke *cluster* 3, yaitu Aceh, Jambi, Bengkulu, Lampung, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, Papua Barat, dan Papua.

Perbandingan Metode *K-means* dan *K-means With Outlier Detection*

Dilakukan perbandingan nilai *R-squared* antara metode *k-means* dan *k-means* with *outlier detection* untuk menentukan metode yang terbaik. Berdasarkan output pada *software* R, diperoleh nilai *R-squared* masing-masing metode adalah sebagai berikut.

Tabel 1. Nilai *R-squared* Metode *K-means* dan *K-means With Outlier Detection*

Metode	Nilai R-squared
k-means	0,3151
k-means with outlier detection	0,5899



Gambar 2. Visualisasi Hasil Pengelompokkan

Berdasarkan perbandingan nilai *R-squared* kedua metode pada Tabel 1, didapatkan hasil bahwa nilai *R-squared* metode *k-means with outlier detection* sebesar 58,99%, nilai ini lebih besar daripada nilai *R-squared* metode *k-means*, yaitu hanya sebesar 31,51%. Hal ini menunjukkan bahwa hasil pengelompokkan metode *k-means with outlier detection* lebih baik daripada metode *k-means*. Oleh karena itu, metode *k-means with outlier detection* adalah metode yang terbaik untuk mengelompokkan provinsi-provinsi di Indonesia dalam kasus ini. Berikut adalah visualisasi hasil pengelompokkan dengan *k-means with outlier detection*.

Berdasarkan visualisasi hasil pengelompokkan pada Gambar 2, terlihat bahwa pengelompokkan dengan *metode* k-means with outlier detection telah mengelompokkan objek berdasarkan jarak terdekatnya dan cukup jelas batas-batas antar-*cluster*-nya. Selain itu, jumlah anggota setiap *cluster* telah terdistribusi sesuai karakteristiknya masing-masing dan tidak terdapat *cluster* yang memiliki jumlah anggota yang sangat minim atau bahkan satu. Hal ini dapat mengindikasikan bahwa *cluster* telah terbentuk dengan cukup baik.

Identifikasi Karakter Cluster

Berdasarkan hasil pengelompokkan dengan metode yang terbaik, yaitu metode *k-means with outlier detection*, untuk melakukan identifikasi karakteristik *cluster*, maka harus memprofilkan hasil *cluster* melalui nilai-nilai *centroid* tiap *cluster*. Berikut hasil perhitungan *centroid* tiap variabel untuk masing-masing *cluster*.

Berdasarkan hasil perhitungan *centroid* masing-masing *cluster* pada Tabel 2, *cluster* 1 memiliki infrastuktur kesehatan dan tenaga medis yang secara umum sangat minim dan merupakan yang paling sedikit dibandingkan *cluster* lain. Hal ini ditunjukkan dari rata-rata rasio puskesmas (V1), rasio rumah sakit (V2), rasio perawat (V6), dan rasio bidan (V7) yang dimiliki oleh anggota-angggota dalam *cluster* 1 merupakan yang terendah dibandingkan 2 *cluster* lainnya. Walaupun rasio klinik (V3) dan rasio dokternya (V5) menengah, serta rasio posyandunya (V4) tertinggi, hal ini masih menunjukkan *cluster* 1 memiliki fasilitas kesehatan yang kurang memadai. Selain itu, *cluster* 1 juga memiliki jumlah kasus terkonfirmasi positif (V8), *fatality rate (V10),* dan jumlah kasus *active (V11)* yang tertinggi dibandingkan 2 *cluster* lain, walaupun *recovery rate*-nya *(V9)* merupakan yang tertinggi juga. Hal ini cukup mengkhawatirkan mengingat fasilitas kesehatan yang tersedia sangat minim sehingga pasien COVID-19 tidak dapat ditangani dengan baik. Oleh karena itu, *cluster* 1 merupakan kelompok provinsi dengan jumlah kasus COVID-19 tinggi, namun fasilitas kesehatan kurang memadai.

Cluster 2 memiliki infrastruktur kesehatan dan tenaga medis yang secara umum sangat memadai dan merupakan yang paling banyak dibandingkan cluster lain. Hal ini terlihat dari rasio rumah sakit (V2), rasio klinik (V3), rasio dokter (V5), dan rasio perawat (V6) yang tertinggi dibandingkan 2 *cluster* lainnya. Walaupun rasio puskesmas (V1) dan rasio bidannya (V7) menengah, serta rasio posyandu (V4) merupakan yang tertinggi, hal ini masih menunjukkan *cluster* 1 memiliki fasilitas kesehatan yang memadai. Akan tetapi, jumlah kasus terkonfirmasi positif (V8) dan jumlah kasus *active*-nya (V11) masih tergolong tinggi, walaupun tidak setinggi di *cluster* 1. Beruntungnya, hal ini dapat ditangani dengan baik mengingat fasilitas kesehatan yang tersedia sangat memadai sehingga *recovery rate* (V9) cukup tinggi dan *fatality rate* (V10) merupakan yang terendah dibandingkan 2 *cluster* lainnya.

Tabel 1. Nilai *R-squared* Metode *K-means* dan *K-means With Outlier Detection*

Cluster	*V1*	*V2*	*V3*	*V4*	*V5*	*V6*	*V7*	*V8*	*V9*	*V10*	*V11*
1	*-0,64*	*-0,45*	*-0,06*	*0,18*	*-0,29*	*-0,67*	*-0,26*	*0,27*	*0,37*	*0,69*	*0,25*
2	*0,06*	*0,88*	*-0,04*	*-0,50*	*0,66*	*0,77*	*-0,26*	*-0,38*	*0,33*	*-0,49*	*-0,48*
3	*0,49*	*-0,34*	*-0,35*	*0,05*	*-0,43*	*-0,00*	*0,63*	*-0,47*	*-0,32*	*-0,15*	*-0,50*

Oleh karena itu, *cluster* 2 merupakan kelompok provinsi dengan jumlah kasus COVID-19 tinggi dan fasilitas kesehatan yang cukup memadai.

Cluster 3 memiliki infrastruktur kesehatan dan tenaga medis yang secara umum tidak terlalu tinggi dan tidak terlalu rendah dibandingkan 2 *cluster* lainnya. Hal ini ditunjukkan dengan rasio puskesmas (V1) dan rasio bidan (V7) yang merupakan tertinggi, tetapi rasio klinik (V3) dan rasio dokternya (V5) merupakan yang terendah. Kemudian, rasio rumah sakit (V2), rasio posyandu (V4), dan rasio perawat (V6) pada *cluster* 3 termasuk menengah. Sedangkan jumlah kasus terkonfirmasi positif (V8), jumlah kasus *active* (V11), dan *recoveray rate cluster* 3 merupakan yang terendah. Untuk *fatality rate* (V10) *cluster* 3 menengah dibandingkan 2 *cluster* lainnya. Oleh karena itu, *cluster* 3 adalah kelompok provinsi dengan jumlah kasus COVID-19 rendah dan fasilitas kesehatan menengah.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dapat diambil kesimpulan sebagai berikut:

1. Untuk melihat metode *clustering* mana yang lebih baik, dapat dilakukan perbandingan nilai *R-squared* antarmetode. Dalam penelitian ini, nilai *R-squared* metode *k-means* (0,3151) lebih kecil dibandingkan metode *k-means with outlier detection (0,5899), sehingga metode k-means with outlier detection menghasilkan clustering yang lebih baik.* Dengan menggunakan metode *clustering k-means with outlier detection*, diperoleh jumlah *cluster* optimum sebanyak 3 *cluster*.
2. *Cluster* 1 memiliki infrastuktur kesehatan dan tenaga medis yang secara umum sangat minim dan merupakan yang paling sedikit. Oleh karena itu, *cluster* 1 merupakan kelompok provinsi dengan jumlah kasus COVID-19 tinggi, namun fasilitas kesehatan kurang memadai. *Cluster* ini terdiri dari Provinsi Sumatera Utara, Sumatera Barat, Riau, Sumatera Selatan, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Barat, Kalimantan Selatan, Kalimantan Timur, dan Sulawesi Selatan.
3. *Cluster* 2 memiliki infrastruktur kesehatan dan tenaga medis yang secara umum sangat memadai dan merupakan yang paling banyak. Oleh karena itu, *cluster* 2 merupakan kelompok provinsi dengan jumlah kasus COVID-19 tinggi dan fasilitas kesehatan yang cukup memadai. *Cluster* ini terdiri dari Provinsi Bangka Belitung, Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Bali, Kalimantan Utara, Sulawesi Utara, dan Maluku.
4. *Cluster* 3 memiliki infrastruktur kesehatan dan tenaga medis yang secara umum tidak terlalu tinggi dan tidak terlalu rendah dibandingkan 2 *cluster* lainnya. Oleh karena itu, *cluster* 3 adalah kelompok provinsi dengan jumlah kasus COVID-19 rendah dan fasilitas kesehatan menengah. *Cluster* ini terdiri dari Provinsi Aceh, Jambi, Bengkulu, Lampung, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, Papua Barat, dan Papua.

Saran

1. Kebijakan pemerintah pada *cluster* 1 hendaknya meningkatkan kualitas dan kuantitas infrastrukur kesehatan dalam penanganan kasus COVID-19 yang cukup tinggi agar lebih maksimal dan mengalokasikan vaksin terlebih dahulu ke *cluster* 1. Untuk *cluster* 2 harus mempertahankan kualitas dan kuantitas infrastruktur kesehatan yang memadai dalam penanganan kasus COVID-19 yang cukup tinggi. Sedangkan, untuk *cluster* 3, kualitas infrastruktur kesehatan harus ditingkatkan agar penanganan kasus COVID-19 lebih maksimal, walaupun saat ini jumlah kasusnya masih tergolong rendah.
2. Dapat dilakukan pengelompokkan menggunakan metode *clustering* lain dan indikator perbandingan antarmetode

lainnya selain *R-squared* serta penambahan *series* data terbaru agar kualitas *cluster* yang dihasilkan lebih baik.

DAFTAR PUSTAKA

Ariawan, Pasek Agus. 2019. Optimasi Pengelompokkan Data Pada Metode - *K-Means* dengan Analisis *Outlier*. Jurnal Nasional Teknologi dan Sistem Informasi, Vol. 5, No. 2.

Charrad, M., dkk.. 2014. NbClust Package for Determining the Best Number of Clusters. https://rdrr.io/cran/NbClust/man/NbClust.html#:~:text=NbClust%20package%20provides%2030%20indices,distance%20measures%2C%20and%20clustering%20methods. (Diakses 13 Desember 2020)

Chawla, Sanjay dan Gionis, Aristides. (2013). k-means--: A unified approach to clustering and outlier detection.

Dwitri, Nayuni, dkk. 2020. Penerapan Algoritma *K-Means* dalam Menentukan Tingkat Penyebaran Pandemi *Covid-19* di Indonesia, Jurnal Teknologi Infomasi, Vol. 4, No. 1, hal. 128-132.

Ikatan Dokter Indonesia (IDI). 2020. Statistik Anggota. http://www.idionline.org/statistik/ (Diakses 12 Desember 2020).

Johnson, Richard A. dan Wichern, Dean W. (2007). *Applied Multivariate Statistical Analysis Sixth Edition*. United States of America : Pearson Education, Inc.

Kementerian Kesehatan. 2019. Rekapitulasi SDM Kesehatan yang didayagunakan di Fasyankes pada tahun 2019. http://bppsdmk.kemkes.go.id/info_sdmk/history/ (Diakses 11 Desember 2020).

Khomarudin, Agus Nur. 2018. Teknik Data Mining : Algoritma *K-means Clustering.* https://ilmukomputer.org/wp-content/uploads/2018/05/agus-k-means-clustering.pdf . (Diakses 13 Desember 2020)

Muhyiddin. 2020. Covid-19 *New Normal* dan Perencanaan Pembangunan di Indonesia. *The Indonesian Journal of Development Planning*, Vol. 4, No. 2.

Noviyanto. 2020. Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia, Jurnal Informatika dan Komputer, Vol. 22, No. 2.

Patel, Maulik. 2016. Clustering Based Outlier Detection Technique. https://rpubs.com/maulikpatel/228345 (Diakses 12 Desember 2020).

Rembulan, Glisina Dwinoor. dkk. 2020. Kebijakan Pemerintah Mengenai *Coronavius Disease* (COVID-19) di Setiap Provinsi di Indonesia Berdasarkan Analisis Klaster, *Jurnal of Industrial Engineering and Management Systems*, Vol. 13, No. 2.

Sindi, Sukma. dkk. 2020. Analisis Algoritma K-Medoids Clustering dalam Pengelompokkan Penyebaran Covid-19 di Indonesia, Jurnal Teknologi Informasi, Vol. 4, No. 1.

Solichin, Achmad dan Khansa Khairunnisa. 2020. Klasterisasi Persebaran Virus Coroana (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means, *Fountain of Informatics Journal*, Vol. 5, No. 2.

Petunjuk Penulisan
JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

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

Naskah dibuat menggunakan *Microsot Office Word* 2010. Seluruh bagian dalam naskah diketik dengan huruf *Times New Roman*, ukuran 12, spasi 1,5, ukuran kertas A4 dan marjin 2 cmuntuk 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 table diletakkan di bagian atas table 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., Brunsdon, 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 Sepetember, 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 Keseimabangan Umum (Computable General Equilibrium Model. *Tesis*. Jakarta.

Artikel di koran, majalah, dan periodik sejenis

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