

JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 15, NOMOR 2, DESEMBER 2023 ISSN 2086 – 4132

Perancangan Kembali Antarmuka Web BPS dengan Pendekatan User Centered Design

CHAIRUNNISA FAUZIA SAMU, LUTFI RAHMATUTI MAGHFIROH

Perbandingan Ordinal Forest dan Regresi Logistik Ordinal
(Studi Kasus: Indeks Desa Membangun Provinsi Maluku Utara 2020)

M. YUNUS, KHAIRIL ANWAR NOTODIPUTRO, BAGUS SARTONO

Unveiling Spatial Disparities: Exploring High-Risk Diarrhea Among Children Under Five
Using Geographically Weighted Quantile Regression

WARA ALFA SYUKRILLA, YUDHIE ANDRIYANA, ANNELEEN VERHASSELT

Peningkatan Kualitas Statistik Resmi Produktivitas Padi Melalui Imputasi Data Non-Respons
Menggunakan Model Aditif Geospasial

MUHLIS ARDIANSYAH

Metode Hutan Ordinal untuk Klasifikasi Desa Sesuai dengan Status Indeks Desa Membangun

DWI AGUSTIN NURIANI SIRODJ, KHAIRIL ANWAR NOTODIPUTRO,
BAGUS SARTONO

Pemodelan *Mixed Geographically Weighted Regression-Spatial Autoregressive* (MGWR-SAR) pada Kasus HIV di Indonesia

ANIK DJURAIDAH, RAHMA ANISA, ARNA RISTIYANTI TARIDA,
MUFTIH ALWI ALIU, CINTIA SEPTEMBERINI, YUFAN PUTRI ASTRINI,
GUSTI TASYA MEILANIA



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.

Editor in Chief:	Rani Nooraeni, SST., M.Stat.
Managing Editor:	Dr. Fitri Kartiasih, SST., S.E., M.Si.
Editor:	Prof. Setia Pramana, S.Si., M.Sc., Ph.D. Dr. Achmad Syahrul Choir, SST., M.Stat. Lutfi Rahmatuti Maghfiroh, SST., M.T.
Assistant Editor dan TI:	Geri Yesa Ermawan, S.Tr.Stat. Muhammad Luqman, S.Tr.Stat.
Mitra Bestari:	Dr. Novianto Budi Kurniawan, SST., M.T. Dr. Azka Ubaidillah Dr. Sarni Maniar Berliana Prof. Dr.rer.pol. Heri Kuswanto, M.Si. Dr. Eng. Arie Wahyu Wijayanto, SST., M.T. Dr. Novi Hidayat Puspongoro, S.Si., M.Stat. Budi Yuniarto, SST., M.Si. Prof. Ir. Sigit Nugroho, M.Sc., Ph.D. Aswi, S.Pd., M.Si. Ph.D.

Alamat Redaksi:

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

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

PENGANTAR REDAKSI

Puji syukur kehadiran Allah, Tuhan Yang Maha Esa, “Jurnal Aplikasi Statistika dan Komputasi Statistik” Volume 15, Nomor 2, Desember 2023 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, Desember 2023

Ketua Dewan Redaksi,

Rani Nooraeni, SST., M.Stat.

JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 15, NOMOR 2, DESEMBER 2023

DAFTAR ISI

Pengantar Redaksi	iii
Daftar Isi	iv
Abstrak	v-xvi
Perancangan Kembali Antarmuka Web BPS dengan Pendekatan User Centered Design <i>Chairunnisa Fauzia Samu, Lutfi Rahmatuti Maghfiroh</i>	1-18
Perbandingan Ordinal Forest dan Regresi Logistik Ordinal (Studi Kasus: Indeks Desa Membangun Provinsi Maluku Utara 2020) <i>M. Yunus, Khairil Anwar Notodiputro, Bagus Sartono</i>	19-30
Unveiling Spatial Disparities: Exploring High-Risk Diarrhea Among Children Under Five Using Geographically Weighted Quantile Regression <i>Wara Alfa Syukrilla, Yudhie Andriyana, Anneleen Verhasselt</i>	31-42
Peningkatan Kualitas Statistik Resmi Produktivitas Padi Melalui Imputasi Data Non-Respons Menggunakan Model Aditif Geospasial <i>Muhlis Ardiansyah</i>	43-52
Metode Hutan Ordinal untuk Klasifikasi Desa Sesuai dengan Status Indeks Desa Membangun <i>Dwi Agustin Nuriani Sirodj, Khairil Anwar Notodiputro, Bagus Sartono</i>	53-64
Pemodelan <i>Mixed Geographically Weighted Regression-Spatial Autoregressive</i> (MGWR-SAR) pada Kasus HIV di Indonesia <i>Anik Djuraidah, Rahma Anisa, Arna Ristiyanti Tarida, Muftih Alwi Aliu, Cintia Septemberini, Yufan Putri Astrini, Gusti Tasya Meilania</i>	65-76

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

DDC: 315.98

Chairunnisa Fauzia Samu, Lutfi Rahmatuti
Maghfiroh

Perancangan Kembali Antarmuka Web
BPS dengan Pendekatan User Centered
Design

Jurnal Aplikasi Statistika & Komputasi
Statistik, Volume 15, Nomor 2, Desember
2023, hal 1 – 18

Abstrak

Pada tahun 2021, BPS menyelenggarakan Survei Kepuasan Pengguna Website (SKPW). Hasil SKPW menunjukkan bahwa aspek pelayanan website BPS secara keseluruhan berada dalam kategori baik, namun masih ada beberapa aspek yang memiliki nilai kepuasan pengguna yang masih kurang dibandingkan dengan harapan pengguna seperti kelengkapan metadata, kesesuaian produk dengan kebutuhan, kemudahan navigasi, kemudahan mencari produk, kehandalan fungsi pencarian, ketertarikan tampilan website, ketepatan penyusunan tata letak, ketersediaan dua bahasa, dan kemudahan mendapatkan panduan penggunaan website. Oleh karena itu, perlu adanya perancangan user interface website BPS yang baru yang dapat memuaskan pengguna website sesuai dengan harapan pengguna. Metode yang digunakan dalam penelitian ini adalah User Centered Design (UCD) dengan menggunakan User Experience Questionnaire (UEQ) sebagai metode evaluasi. Selain itu, dilakukan juga evaluasi kepuasan pengguna terhadap rancangan antarmuka yang baru untuk dibandingkan dengan SKPW 2021. Agar ukuran kepuasan pengguna yang dihasilkan dapat dibandingkan dengan hasil evaluasi SKPW 2021, digunakan metode analisis kepuasan

yang sama dengan yang digunakan di SKPW 2021 yaitu Importance Performance Analysis (IPA). User interface baru yang telah dirancang telah dievaluasi ke pengguna sebanyak dua kali iterasi dengan hasil evaluasi terakhir yaitu tingkat kepuasan seluruh aspek pelayanan yang menjadi cakupan penelitian ini telah mengalami peningkatan dibandingkan dengan hasil SKPW 2021. User interface baru sudah mempunyai tingkat kepuasan yang lebih baik dibandingkan dengan hasil SKPW 2021 dan mempunyai hasil penilaian UEQ berupa Good pada 3 aspek (daya tarik, stimulasi, dan kebaruan) dan Above Average pada 3 aspek (kejelasan, efisiensi, dan ketepatan).

Kata kunci: user interface, website, user centered design, user experience questionnaire, importance performance analysis

DDC: 315.98

M. Yunus, Khairil Anwar Notodiputro,
Bagus Sartono

Perbandingan Ordinal Forest dan Regresi
Logistik Ordinal (Studi Kasus: Indeks Desa
Membangun Provinsi Maluku Utara 2020)

Jurnal Aplikasi Statistika & Komputasi
Statistik, Volume 15, Nomor 2, Desember
2023, hal 19 – 30

Abstrak

Terdapat banyak metode yang digunakan untuk mengidentifikasi prediktor penting dan memprediksi nilai dari peubah respons ordinal. Namun, terdapat metode prediksi untuk peubah respons ordinal yang tidak

menggunakan sifat ordinal tersebut. Penelitian ini menggunakan metode ordinal forest dan metode regresi logistik ordinal. Nilai akurasi dan kappa metode ordinal forest pada setiap ulangan (5 ulangan) selalu lebih besar dari pada regresi logistik ordinal. Selanjutnya, nilai akurasi dan kappa setiap kelompok berdasarkan PDRB pada metode ordinal forest selalu lebih besar dari pada regresi logistik ordinal. Sehingga metode ordinal forest lebih baik digunakan pada data peringkat status indeks desa membangun Provinsi Maluku Utara 2020.

Kata kunci: indeks desa membangun, ordinal forest, peubah respons ordinal, regresi logistik ordinal

DDC: 315.98

Wara Alfa Syukrilla, Yudhie Andriyana,
Anneleen Verhasselt

Unveiling Spatial Disparities: Exploring
High-Risk Diarrhea Among Children Under
Five Using Geographically Weighted
Quantile Regression

Jurnal Aplikasi Statistika & Komputasi
Statistik, Volume 15, Nomor 2, Desember
2023, hal 31 – 42

Abstrak

Penelitian ini menyelidiki pengaruh persentase akses air bersih, persentase kebiasaan mencuci tangan, dan kategorisasi toilet sehat pada kuantil atas risiko diare balita di Kota Bandung, Indonesia, menggunakan model Geographically Weighted Quantile Regression pada persentil ke-75 ($\tau = 0,75$). Bandwidth optimal dipilih menggunakan validasi silang. Hasil penelitian menunjukkan bahwa signifikansi, kekuatan, dan arah hubungan antara diare dan faktor risikonya tergantung pada lokasinya. Pada kuantil

atas $\tau = 0,75$ Kecamatan Panyileukan diprediksi memiliki risiko diare tertinggi. Di kabupaten ini, ketiga prediktor berpengaruh signifikan terhadap risiko diare pada balita, dengan variabel persentase rumah yang mempraktikkan kebiasaan cuci tangan adalah variabel paling besar pengaruhnya dalam menurunkan risiko diare. Kesimpulannya, akses air bersih, kebiasaan cuci tangan, dan kategori toilet merupakan faktor risiko potensial terjadinya diare pada anak risiko tinggi. Metode GWQR memungkinkan pembuat keputusan untuk menangani masalah diare dengan tepat berdasarkan prediktor mana yang memiliki pengaruh besar pada daerah tertentu yang diminati. Selain itu, GWQR dapat digunakan untuk menyelidiki efek dari berbagai strategi intervensi dan secara efektif mengalokasikan sumber daya terbatas yang tersedia sesuai lokasi yang paling membutuhkannya.

Kata kunci: *Geographically Weighted Quantile Regression*, balita, diare

DDC: 315.98

Muhlis Ardiansyah

Peningkatan Kualitas Statistik Resmi Produktivitas Padi Melalui Imputasi Data Non-Respons Menggunakan Model Aditif Geospasial

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 2, Desember 2023, hal 43 – 52

Abstrak

Penelitian ini dilatarbelakangi oleh masalah non-respons pada Survei Ubinan yang diselenggarakan oleh Badan Pusat Statistik (BPS) sebagai penyedia statistik resmi. BPS memiliki visi sebagai penyedia data statistik berkualitas untuk Indonesia maju. Penanganan non-respons sangat penting untuk mendukung visi tersebut karena non-respons berpotensi menyebabkan beberapa karakteristik sampel menjadi tidak terwakili. Penelitian ini mengusulkan teknik imputasi data non-

respon melalui pemodelan statistik. Model yang diusulkan adalah model aditif dengan penambahan fungsi pemulusan geospasial thin plate regression splines (TP) dan Gaussian process (GP). Pemilihan model terbaik berdasarkan Mean Squared Error of Prediction (MSEP) terkecil dari 1000 iterasi. Kemudian dibandingkan rata-rata produktivitas padi antara teknik mengabaikan non-respons (listwise deletion) dan imputasi melalui tiga skenario data non-respons. Hasil penelitian menunjukkan bahwa model dengan penambahan fungsi pemulus GP memberikan kinerja terbaik dengan MSEP tekecil. Hasil lainnya menunjukkan bahwa metode imputasi data non-respons lebih baik dibandingkan dengan mengabaikan non-respons. BPS dapat mempertimbangkan metode imputasi untuk meningkatkan kualitas data statistik resmi produktivitas padi.

Kata kunci: imputasi, geospasial, non-respons, statistik resmi

DDC: 315.98

Dwi Agustin Nuriani Sirodj, Khairil Anwar Notodiputro, Bagus Sartono

Metode Hutan Ordinal untuk Klasifikasi Desa Sesuai dengan Status Indeks Desa Membangun

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 2, Desember 2023, hal 53 – 64

Abstrak

Metode hutan acak merupakan metode klasifikasi berbasis pohon yang cukup populer. Metode ini jika diterapkan pada peubah respons ordinal akan memberikan hasil yang mengubah karakteristik dari data ordinal itu sendiri. Dalam tulisan ini akan dibahas kinerja dari metode hutan ordinal dan hutan ordinal Naïve untuk

klasifikasi status berbagai desa sesuai dengan Indeks Desa Membangun (IDM) di kabupaten Tasikmalaya dan Indramayu. Hasilnya memperlihatkan bahwa untuk Kabupaten Tasikmalaya kinerja metode hutan ordinal mengungguli kinerja hutan ordinal Naïve dengan rata-rata akurasi sebesar 73.8% dan rata-rata nilai kappa sebesar 0.18 sedangkan pada Kabupaten Indramayu kinerja metode hutan ordinal Naïve yang justru mengungguli metode hutan ordinal dengan rata-rata akurasi sebesar 51.6 % dan rata-rata nilai kappa sebesar 0.162. Selain itu ditunjukkan pula bahwa peubah yang penting dalam proses klasifikasi status IDM di Kabupaten Tasikmalaya dan Kabupaten Indramayu adalah peubah Jumlah Koperasi dan Jarak terdekat ke rumah sakit.

Kata kunci: Hutan ordinal, Naïve Hutan ordinal, Klasifikasi Status IDM

DDC: 315.98

Anik Djuraidah, Rahma Anisa, Arna Ristiyanti Tarida, Muftih Alwi Aliu, Cintia Septemberini, Yufan Putri Astrini, Gusti Tasya Meilania

Pemodelan *Mixed Geographically Weighted Regression-Spatial Autoregressive* (MGWR-SAR) pada Kasus HIV di Indonesia

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 2, Desember 2023, hal 67– 80

Abstrak

Pada umumnya regresi spasial memodelkan salah satu efek spasial, yaitu dependensi atau heterogenitas spasial. Untuk efek dependensi spasial, model yang telah sering dipakai mengikuti taksonomi Elhast, dengan dependensi spasial terdapat pada peubah respon, prediktor, atau galat. Sedangkan untuk efek heterogenitas spasial umumnya menggunakan model regresi terboboti geografis (GWR) atau bila terdapat prediktor yang global digunakan regresi terboboti geografis campuran

(MGWR). Data yang digunakan pada penelitian ini adalah kasus Human Immunodeficiency Virus (HIV) per 100.000 penduduk sebagai peubah respon, dan populasi kunci, kasus positif pada ibu hamil, pasien tuberkulosis, tingkat kemiskinan, tingkat pengangguran sebagai prediktor. Pada data yang digunakan terdapat efek dependensi dan heterogenitas spasial. MGWR-SAR merupakan model yang dapat digunakan bila pada data terdapat kedua efek spasial tersebut. Penelitian ini bertujuan untuk menentukan faktor-faktor yang mempengaruhi kasus HIV di kabupaten/kota di Indonesia menggunakan model spasial. Hasil penelitian menunjukkan bahwa model MGWR-SAR $(0, k_c, k_v)$ merupakan model terbaik yang terpilih berdasarkan nilai RMSE terkecil yaitu sebesar 0,54066. Prediktor populasi kunci memiliki pengaruh global dan signifikan. Prediktor lain memiliki pengaruh secara lokal adalah kasus positif pada ibu hamil, pasien tuberkulosis, tingkat kemiskinan, dan tingkat pengangguran.

Kata kunci: GWR, HIV, MGWR, MGWR-SAR, SAR

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

DDC: 315.98

Chairunnisa Fauzia Samu, Lutfi Rahmatuti
Maghfiroh

Perancangan Kembali Antarmuka Web
BPS dengan Pendekatan User Centered
Design

Jurnal Aplikasi Statistika & Komputasi
Statistik, Volume 15, Nomor 2, Desember
2023, hal 1 – 18

Abstract

In 2021, BPS conducted a Website User Satisfaction Survey (SKPW). The results of the SKPW show that the overall service aspect of the BPS website is in the good category, but there are still some aspects that have user satisfaction values that are still lacking compared to user expectations such as completeness of metadata, suitability of products with needs, ease of navigation, ease of finding products, reliability of functions. search, attractiveness of website appearance, accuracy of layout preparation, availability of two languages, and ease of obtaining website usage guidelines. Therefore, it is necessary to design a new BPS website user interface that can satisfy website users in accordance with user expectations. The

method used in this study is User Centered Design (UCD) using the User Experience Questionnaire (UEQ) as an evaluation method. In addition, an evaluation of user satisfaction with the new interface design was also carried out to be compared with SKPW 2021. In order for the resulting measure of user satisfaction to be compared with the results of the evaluation of SKPW 2021, the same satisfaction analysis method was used as used in SKPW 2021, namely Importance Performance Analysis (IPA). The new user interface that has been designed has been evaluated to the user for two iterations with the final evaluation result, namely the level of satisfaction of all service aspects that are the scope of this research has increased compared to the results of SKPW 2021. The new user interface already has a better level of satisfaction compared to the results of SKPW 2021 and has a UEQ assessment result of Good in 3 aspects (attractiveness, stimulation, and novelty) and Above Average in 3 aspects (clarity, efficiency, and accuracy)

Keywords: user interface, website, user centered design, user experience questionnaire, importance performance analysis

DDC: 315.98

M. Yunus, Khairil Anwar Notodiputro,
Bagus Sartono

Perbandingan Ordinal Forest dan Regresi
Logistik Ordinal (Studi Kasus: Indeks Desa
Membangun Provinsi Maluku Utara 2020)

Jurnal Aplikasi Statistika & Komputasi
Statistik, Volume 15, Nomor 2, Desember
2023, hal 19 – 30

Abstract

There are many methods used to identify the important predictors and predict the value of the ordinal response variable. However, there are prediction methods for ordinal response variables that do not use the

ordinal nature. This study uses the ordinal forest method and the ordinal logistic regression method. The accuracy and kappa value of the ordinal forest method in each replication (5 replications) is always greater than the ordinal logistic regression. Furthermore, the value of accuracy and kappa of each group based on GRDP in the ordinal forest method is always greater than the ordinal logistic regression. So that the ordinal forest method is better used in data on village development index in North Maluku Province 2020.

Keywords: ordinal forest, ordinal logistic regression, ordinal response variable, village development index

DDC: 315.98

Wara Alfa Syukrilla, Yudhie Andriyana,
Anneleen Verhasselt

Unveiling Spatial Disparities: Exploring
High-Risk Diarrhea Among Children
Under Five Using Geographically
Weighted Quantile Regression

Jurnal Aplikasi Statistika & Komputasi
Statistik, Volume 15, Nomor 2, Desember
2023, hal 31 – 42

Abstract

We investigate the impact of the percentage of clean water access, the percentage of handwashing habits, and the toilet category factors on the upper quantile of toddlers' diarrhea risks in Bandung City, Indonesia, using the Geographically Weighted Quantile Regression model on the 75th percentile ($\tau=0.75$). The optimum bandwidth was selected using cross-validation. The results show that the significance, strength, and direction of the

relationship between diarrhea and its risk factors depend on the location. At the upper quantile $\tau = 0.75$, the Panyileukan district is predicted to have the highest diarrhea risk. In this district, all three predictors significantly affect the toddlers' diarrhea risk, with the variable of the percentage of houses practicing hand washing habits observed to reduce diarrhea risk the most. In conclusion, clean water access, handwashing habits, and toilet category are the potential risk factors for high-risk childhood diarrhea. This method is powerful as it would allow the decision-maker to handle the diarrhea problem aptly by focusing on the predictor that has a significant impact on a particular district of interest. And it can be used to investigate the effect of various intervention strategies and effectively allocate the limited available resources according to the most important locations.

Keywords: Geographically Weighted Quantile Regression, toddlers, diarrhea

DDC: 315.98

Muhlis Ardiansyah

Peningkatan Kualitas Statistik Resmi Produktivitas Padi Melalui Imputasi Data Non-Respons Menggunakan Model Aditif Geospasial

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 2, Desember 2023, hal 43 – 52

Abstract

This study is motivated by the non-response problem in the Crop Cutting Survey conducted by the BPS-Statistics Indonesia as the official statistics provider. BPS has a vision of providing quality statistical data for advanced Indonesia. Handling non-response is essential to supporting this vision because non-response can potentially cause some sample characteristics to be unrepresented. This study proposed a non-response data

imputation technique through statistical modeling. The proposed model was an additive model with the addition of geospatial smoothing functions of thin plate regression splines (TP) and Gaussian process (GP). Selection of the best model based on the smallest MSE of 1000 iterations. Then we compared the average rice productivity between listwise deletion and imputation techniques through three scenarios of non-response data. The results showed that the model with the addition of the GP smoothing function gave the best performance with the smallest MSE. The other results showed that the imputation method of non-response data is better than ignoring non-response. BPS can consider the imputation method to improve the quality of official statistics on rice productivity.

Keywords: imputation, geospatial, non-response, official statistics

DDC: 315.98

Dwi Agustin Nuriani Sirodj, Khairil Anwar Notodiputro, Bagus Sartono

Metode Hutan Ordinal untuk Klasifikasi Desa Sesuai dengan Status Indeks Desa Membangun

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 2, Desember 2023, hal 53 – 64

Abstract

The random forest (RF) method is a popular tree-based classification method. However, when faced with ordinal response variables, the prediction results of the RF method cannot maintain the characteristics of the ordinal data itself. So, in this paper discusses the performance

of the ordinal forest and Naïve ordinal forest methods for the rural's classification based on the Village Development Index (IDM) in Tasikmalaya and Indramayu. The results show that in the Tasikmalaya district, the OF method performs better, with an average accuracy of 73.8% and an average kappa value of 0.18. However, for the Indramayu district, the Naïve OF method looks better than the OF method, producing an average accuracy of 51.6% and an average kappa value of 0.162. In addition, it is also shown that the important variables in the rural's classification in Tasikmalaya and Indramayu districts are the variables of the number of cooperatives and the closest distance to the hospital.

Keywords: Ordinal Forest, Naïve Ordinal Forest, rural's classification

DDC: 315.98

Anik Djuraidah, Rahma Anisa, Arna Ristiyanti Tarida, Muftih Alwi Aliu, Cintia Septemberini, Yufan Putri Astrini, Gusti Tasya Meilania

Pemodelan *Mixed Geographically Weighted Regression-Spatial Autoregressive* (MGWR-SAR) pada Kasus HIV di Indonesia

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 2, Desember 2023, hal 67– 80

Abstract

In general, spatial regression is used to model one of the spatial effects, namely spatial dependency or heterogeneity. For the effects of spatial dependencies, the models that have been used frequently follow Elhost's taxonomy, with the spatial dependencies being on the response, predictor, or error. Whereas for the effect of spatial heterogeneity generally use geographically weighted regression models (GWR) or if there are global predictors use mixed geographically weighted regression

(MGWR). The data used in this study are cases of Human Immunodeficiency Virus (HIV) per 100,000 population as a response variable, and key populations, positive cases in pregnant women, tuberculosis patients, poverty rate, and unemployment rate as predictors. In the data used, there are spatial dependencies and heterogeneity. The MGWR-SAR is a model that can be used if the data has both spatial effects. This study aims to determine the factors influencing HIV cases in districts/cities in Indonesia using a spatial model. The results showed that the combined model of GWR and spatial autoregressive regression (SAR) was the best model. The research results indicate that the MGWR-SAR model (0,k_c,k_v) is the best model selected based on the smallest RMSE value, which is 0.54066. Key population explanatory variables have a global and significant influence. Other explanatory variables that have local influence are positive cases in pregnant women, tuberculosis patients, poverty rates, and unemployment rates.

Keywords: GWR, HIV, MGWR, MGWR-SAR, SAR

PERANCANGAN KEMBALI ANTARMUKA WEB BPS DENGAN PENDEKATAN USER CENTERED DESIGN

Chairunnisa Fauzia Samu¹, Lutfi Rahmatuti Maghfiroh¹

¹ Politeknik Statistika STIS
e-mail: lutfirm@stis.ac.id

Received: November 15, 2022; Revised: July 25, 2023; Accepted: September 18, 2023; Published: December 31, 2023

Abstrak

Pada tahun 2021, BPS menyelenggarakan Survei Kepuasan Pengguna Website (SKPW). Hasil SKPW menunjukkan bahwa aspek pelayanan website BPS secara keseluruhan berada dalam kategori baik, namun masih ada beberapa aspek yang memiliki nilai kepuasan pengguna yang masih kurang dibandingkan dengan harapan pengguna seperti kelengkapan metadata, kesesuaian produk dengan kebutuhan, kemudahan navigasi, kemudahan mencari produk, kehandalan fungsi pencarian, ketertarikan tampilan website, ketepatan penyusunan tata letak, ketersediaan dua bahasa, dan kemudahan mendapatkan panduan penggunaan website. Oleh karena itu, perlu adanya perancangan user interface website BPS yang baru yang dapat memuaskan pengguna website sesuai dengan harapan pengguna. Metode yang digunakan dalam penelitian ini adalah User Centered Design (UCD) dengan menggunakan User Experience Questionnaire (UEQ) sebagai metode evaluasi. Selain itu, dilakukan juga evaluasi kepuasan pengguna terhadap rancangan antarmuka yang baru untuk dibandingkan dengan SKPW 2021. Agar ukuran kepuasan pengguna yang dihasilkan dapat dibandingkan dengan hasil evaluasi SKPW 2021, digunakan metode analisis kepuasan yang sama dengan yang digunakan di SKPW 2021 yaitu Importance Performance Analysis (IPA). User interface baru yang telah dirancang telah dievaluasi ke pengguna sebanyak dua kali iterasi dengan hasil evaluasi terakhir yaitu tingkat kepuasan seluruh aspek pelayanan yang menjadi cakupan penelitian ini telah mengalami peningkatan dibandingkan dengan hasil SKPW 2021. User interface baru sudah mempunyai tingkat kepuasan yang lebih baik dibandingkan dengan hasil SKPW 2021 dan mempunyai hasil penilaian UEQ berupa Good pada 3 aspek (daya tarik, stimulasi, dan kebaruan) dan Above Average pada 3 aspek (kejelasan, efisiensi, dan ketepatan).

Kata kunci: user interface, website, user centered design, user experience questionnaire, importance performance analysis

Abstract

In 2021, BPS conducted a Website User Satisfaction Survey (SKPW). The results of the SKPW show that the overall service aspect of the BPS website is in the good category, but there are still some aspects that have user satisfaction values that are still lacking compared to user expectations such as completeness of metadata, suitability of products with needs, ease of navigation, ease of finding products, reliability of functions. search, attractiveness of website appearance, accuracy of layout preparation, availability of two languages, and ease of obtaining website usage guidelines. Therefore, it is necessary to design a new BPS website user interface that can satisfy website users in accordance with user expectations. The method used in this study is User Centered Design (UCD) using the User Experience Questionnaire (UEQ) as an evaluation method. In addition, an evaluation of user satisfaction with the new interface design was also carried out to be compared with SKPW 2021. In order for the resulting measure of user satisfaction to be compared with the results of the evaluation of SKPW 2021, the same satisfaction analysis method was used as used in SKPW 2021, namely Importance Performance Analysis (IPA). The new user interface that has been designed has been evaluated to the user for two iterations with the final evaluation result, namely the level of satisfaction of all service aspects that are the scope of this research has increased compared to the results of SKPW 2021. The new user interface already has a better level of satisfaction compared to the results of SKPW 2021 and has a UEQ assessment result of Good in 3 aspects (attractiveness, stimulation, and novelty) and Above Average in 3 aspects (clarity, efficiency, and accuracy)

Keywords: user interface, website, user centered design, user experience questionnaire, importance performance analysis

PENDAHULUAN

Badan Pusat Statistik (BPS) merupakan Lembaga Pemerintah Non Kementrian yang salah satu perannya adalah sebagai penyedia data bagi pemerintah dan masyarakat (BPS, 2021). Dalam mendiseminasikan data statistik yang telah dikumpulkan, salah satu media yang dikembangkan oleh BPS adalah website BPS yang dapat diakses pada alamat www.bps.go.id. Website ini berfungsi sebagai gerbang utama pelayanan statistik untuk publik, baik itu pemerintah maupun masyarakat, terutama bagi yang membutuhkan data dan informasi yang tersedia di website BPS, seperti publikasi data, statistik, maupun indikator yang dihasilkan dari sensus/survei. Website BPS memiliki berbagai macam fitur mulai dari fitur pencarian hingga fitur tabel dinamis yang memungkinkan pengguna memperoleh data yang bisa dikustomisasi variabelnya secara manual. Website BPS juga menyajikan sejumlah produk seperti Publikasi, Berita Resmi Statistik (BRS), Infografis, dan Tabel Statistik.

Dalam rangka meningkatkan kualitas dan kepuasan masyarakat terhadap website BPS, BPS menyelenggarakan Survei Kepuasan Pengguna Website (SKPW) BPS pada rentang waktu 1-30 September 2021. Survei ini bertujuan untuk mengetahui persepsi serta pengalaman pengguna data selama mengakses website BPS, yang kemudian akan dijadikan bahan evaluasi untuk meningkatkan kualitas layanan yang ada di website BPS (BPS, 2021). SKPW 2021 menggunakan rancangan non-probability sampling dengan kuesioner yang disebarakan melalui website dan media sosial BPS serta secara berantai (snowball). Untuk analisis kepuasan pengguna terhadap website BPS, SKPW 2021 menggunakan metode Importance Performance Analysis atau IPA (BPS, 2021).

Hasil evaluasi SKPW 2021 menunjukkan bahwa website BPS secara keseluruhan berada dalam kategori baik, namun masih ada beberapa atribut layanan yang kinerjanya belum memenuhi ekspektasi pengguna, antara lain

kelengkapan metadata, kesesuaian produk yang ada pada website BPS dengan kebutuhan, kemudahan navigasi, kemudahan dalam memperoleh/mencari produk yang dibutuhkan, dan kehandalan fungsi pencarian. Selain itu, aspek pelayanan yang memiliki ekspektasi dan kinerja di bawah rata-rata adalah ketertarikan tampilan website, ketepatan penyusunan tata letak, ketersediaan dua bahasa, dan kemudahan mendapatkan panduan penggunaan website (BPS, 2021). Berdasarkan hasil evaluasi SKPW 2021, dapat ditarik kesimpulan bahwa perlu adanya perbaikan pada user interface website BPS yang baru yang dapat memuaskan pengguna website sesuai dengan harapan pengguna. Untuk dapat memenuhi harapan pengguna, user interface harus dirancang dengan baik sehingga dapat memudahkan pengguna (Johnson, 2008) (Galitz, 2007).

Salah satu metode yang dapat digunakan dalam proses perancangan user interface adalah metode User Centered Design (UCD). Metode ini dipilih karena dalam pendekatan UCD, pengguna dilibatkan selama proses desain sehingga dapat menghasilkan produk yang memenuhi kebutuhan pengguna (Interaction Design Foundation, 2017). Agar perancangan antarmuka website lebih objektif dan konsisten, perlu digunakan pedoman yang dapat menjadi guideline dalam proses perancangan kembali antarmuka website BPS. Salah satu guideline yang dapat digunakan adalah buku *Designing User Interfaces*, sebuah buku pedoman desain user interface yang diterbitkan oleh HYPE4. Guideline yang disajikan dalam buku tersebut telah melalui kajian internal dan eksternal yang ekstensif yang telah dilakukan oleh HYPE4 sejak tahun 2013 (Malewicz & Malewicz, 2020). Dalam metode UCD, salah satu tahapannya adalah mengevaluasi rancangan yang dihasilkan. Pada tahapan evaluasi ini akan digunakan metode evaluasi User Experience Questionnaire (UEQ). UEQ merupakan salah satu metode usability testing dalam bentuk kuesioner yang digunakan untuk mengukur kesan subjektif

pengguna ketika menggunakan suatu produk (Laugwitz, Schrepp, & Held, 2008). Metode evaluasi dengan UEQ dipilih karena kuesioner UEQ cepat dan mudah untuk mengukur pengalaman pengguna ketika menggunakan suatu produk, dalam hal ini antarmuka website BPS yang baru yang akan dilakukan perancangannya di penelitian ini. Selain itu, akan dilakukan juga evaluasi kepuasan pengguna terhadap rancangan antarmuka yang baru. Agar ukuran kepuasan pengguna yang dihasilkan dapat dibandingkan dengan hasil evaluasi SKPW 2021, maka akan digunakan metode analisis kepuasan yang sama dengan yang digunakan di SKPW 2021 yaitu Importance Performance Analysis (IPA). IPA adalah teknik analisis yang berguna untuk mengukur kinerja atribut produk atau layanan berdasarkan kepuasan pengguna sehingga akan teridentifikasi atribut mana saja yang perlu ditingkatkan beserta prioritasnya (Martilla & James, 1977). Oleh karena itu, metode yang akan digunakan untuk evaluasi penelitian ini adalah UEQ dan IPA.

Oleh karena itu dapat disimpulkan bahwa tujuan utama penelitian ini antara lain merancang kembali user interface web BPS dari sisi kemudahan navigasi, kemudahan mencari produk, ketertarikan tampilan website, dan ketepatan penyusunan tata letak dengan menggunakan UCD. Untuk evaluasi produk, peneliti

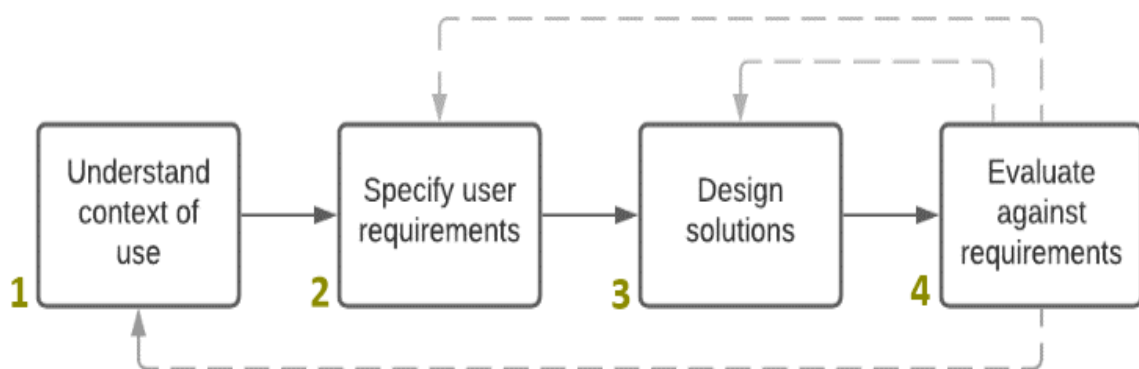
penelitian ini adalah prototype tampilan website BPS yang baru dan pedoman UI.

METODE PENELITIAN

1. Metode

User Interface (UI) Design atau desain antarmuka pengguna adalah proses yang digunakan desainer untuk membangun antarmuka dalam perangkat lunak atau perangkat terkomputerisasi, dengan berfokus pada tampilan atau gaya (Interaction Design Foundation, 2016). Penelitian ini akan dilakukan dengan menggunakan metode User Centered Design (UCD). Metode UCD merupakan suatu proses iteratif yang pada setiap tahap perancangan dan pengembangannya berfokus untuk memahami pengguna serta kebutuhan pengguna dalam menggunakan suatu produk (Interaction Design Foundation, 2017). Setiap iterasi dalam pendekatan UCD melibatkan empat tahap yang berbeda yang dapat dilihat Gambar 1.

Berdasarkan Gambar 1, tahap pertama adalah memahami konteks penggunaan dan mengidentifikasi kebutuhan pengguna. Tahap pertama ini peneliti lakukan dengan menganalisis hasil SKPW. Hasil analisis tersebut mencakup bagaimana pengguna mengakses website BPS dan produk apa saja yang paling sering diakses sehingga informasi yang diperoleh dari hasil analisis tersebut dapat digunakan



Sumber: Interaction Design Foundation (2017)

Gambar 1. Diagram metode User Centered Design

menggunakan UEQ dan IPA. Hasil dari

dalam dua tahapan pertama dari metode UCD. Tahap selanjutnya dari metode UCD

adalah pengembangan solusi desain. Dalam proses perancangan solusi desain akan dilakukan dalam beberapa tahapan antara lain pembuatan wireframe, mockup, dan prototype (Karnawan, Andryanam dan Komalasari 2021). Tahapan selanjutnya adalah mengevaluasi rancangan yang dihasilkan. Ada dua alat evaluasi yang akan digunakan pada tahapan evaluasi yaitu User Experience Questionnaire (UEQ) dan Importance Performance Analysis (IPA).

UEQ merupakan salah satu metode usability testing dalam bentuk kuesioner yang digunakan untuk mengukur kesan subjektif pengguna ketika menggunakan suatu produk (Laugwitz, Schrepp, & Held, 2008). Metode evaluasi dengan UEQ dipilih karena karena kuesioner UEQ cepat dan mudah untuk mengukur pengalaman pengguna ketika menggunakan suatu produk, dalam hal ini antarmuka website BPS yang baru yang akan dilakukan perancangannya di penelitian ini.

IPA adalah teknik analisis yang berguna untuk mengukur kinerja atribut produk atau layanan berdasarkan kepuasan pengguna sehingga akan teridentifikasi atribut mana saja yang perlu ditingkatkan beserta prioritasnya (Martilla & James, 1977). Analisis dengan metode IPA memiliki sejumlah kelebihan dalam mengevaluasi persepsi pengguna terhadap produk, yaitu dapat mengidentifikasi informasi penting, mulai dari aspek apa saja yang harus lebih diperhatikan karena kinerjanya masih kurang dari ekspektasi pengguna, aspek yang kinerjanya sudah sesuai dengan ekspektasi pengguna, hingga aspek yang telah melebihi ekspektasi pengguna (Martilla & James, 1977). IPA digunakan agar ukuran kepuasan pengguna yang dihasilkan dapat dibandingkan dengan hasil evaluasi SKPW 2021. Model yang digunakan dalam IPA adalah model kuadran. Model ini terdiri dari sepasang koordinat dan empat bidang yang berisi elemen kepentingan dan kinerja untuk layanan/produk tertentu, yang mana nilai rata-ratanya dihitung dari penilaian langsung pengguna layanan/produk (Ormanovic et al., 2017).

2. Data dan Sumber Data

Data yang digunakan adalah data tingkat kepentingan dan tingkat kepuasan pengguna yang merupakan data primer dengan pengguna website BPS sebagai sumber data. Untuk memperoleh data evaluasi terkait persepsi pengguna terhadap rancangan tampilan baru website BPS, dilakukan survei penelitian secara online dengan teknik pengumpulan data menggunakan kuesioner. Sumber data yang dikumpulkan dalam penelitian ini berupa data primer yang berasal dari pengguna website BPS. Teknik sampling yang digunakan adalah metode non-probability sampling yaitu teknik pengambilan sampel yang masing-masing anggota populasinya tidak memiliki peluang yang sama untuk terpilih menjadi. Metode sampling ini dipilih karena tidak tersedianya kerangka sampel pengguna website BPS sehingga tidak memungkinkan untuk dilakukan sampling dengan metode probability. Jenis non-probability sampling yang digunakan dalam penelitian ini adalah convenience sampling yaitu teknik penentuan sampel dari orang-orang yang mudah diakses dan cocok sebagai sumber data yang akan berkontribusi pada survei secara sukarela (Sugiyono, 2013).

Pada survei tersebut juga dilakukan usability testing yaitu teknik untuk menguji seberapa mudah sebuah desain digunakan dengan melibatkan sekelompok pengguna yang representatif sebagai peserta pengujian (Moran, 2019). Jenis usability testing yang dilakukan adalah usability testing kuantitatif secara jarak jauh dan unmoderated. Usability testing secara kuantitatif dipilih karena kuesioner UEQ yang akan digunakan berfokus pada pengumpulan angka yang mewakili aspek-aspek pengalaman pengguna. Pendekatan secara jarak jauh dipilih karena kondisi saat ini tidak memungkinkan untuk melakukan pertemuan tatap muka dengan pengguna website BPS yang tersebar di seluruh Indonesia. Sedangkan pendekatan secara unmoderated dipilih agar lebih efisien dari segi waktu dan lebih fleksibel karena tidak perlu melakukan penjadwalan dengan responden, kuesioner dapat diisi oleh responden sesuai waktu luang masing-

masing dan bahkan dapat dilakukan secara bersamaan di satu waktu yang sama oleh lebih dari satu responden sehingga lebih efisien dari segi waktu. Metrik yang diukur dalam survei penelitian ini adalah metrik kontinu berupa penilaian tingkat kepentingan dan kepuasan. Untuk quantitative usability testing yang menggunakan metrik kontinu adalah minimal sebanyak 47 responden (Budiu & Moran, 2021).

HASIL DAN PEMBAHASAN

Analisis Konteks Penggunaan

Sesuai dengan alur perancangan dengan menggunakan UCD, tahap pertama yang perlu dilakukan adalah memahami konteks penggunaan yang didapatkan dari SKPW 2021. Dalam hasil analisis SKPW 2021 terdapat analisis terkait profil pengguna website BPS yaitu meliputi jenis perangkat yang digunakan ketika mengakses website BPS, website BPS yang paling sering dikunjungi, banyaknya pengguna menurut produk yang diakses melalui website BPS, banyaknya pengguna menurut produk yang diunduh melalui website BPS, dan lain-lain.

Dari hasil analisis SKPW 2021 diperoleh informasi bahwa 67,40% pengguna website BPS menggunakan perangkat desktop computer atau laptop untuk mengakses website BPS, 30,51% pengguna menggunakan perangkat gawai atau smartpone, dan 2,09% pengguna menggunakan perangkat tablet. Lebih dari sebagian pengguna memakai desktop/laptop untuk mengakses website BPS. Banyaknya peminat website versi desktop ini menunjukkan bahwa website BPS dirasa lebih nyaman diakses dibandingkan website versi mobile (BPS, 2021).

Berdasarkan produk di website BPS yang paling sering diakses, 27,57% pengguna mengakses publikasi, 21,88% mengakses Berita Resmi Statistik (BRS), 18,14% pengguna mengakses infografis, 11,04% pengguna mengakses metadata, 10,77% pengguna mengakses tabel statistik, 5,38% pengguna mengakses indikator

strategis, dan 5,22% mengakses produk lainnya. Dari hasil analisis ini diperoleh informasi bahwa produk di website BPS yang paling sering diakses adalah produk dengan persentase terbesar yaitu publikasi, sedangkan persentase terendah adalah akses pengguna terhadap indikator strategis (BPS, 2021).

Hasil analisis berdasarkan produk yang diunduh di website BPS menunjukkan bahwa pengguna yang mengunduh publikasi memiliki persentase terbesar, yaitu sebanyak 40,23% dibandingkan dengan produk website BPS lainnya, seperti Berita Resmi Statistik (BRS) sebesar 22,58%, tabel statistik sebesar 20,92%, sedangkan pengguna yang mengunduh infografis hanya sebesar 9,74% (BPS, 2021).

Dari analisis konteks penggunaan ini, diperoleh informasi produk apa saja yang paling sering diakses oleh pengguna di website BPS. Dari informasi tersebut kemudian peneliti membatasi cakupan halaman yang akan di-redesign. Halaman yang menjadi cakupan tersebut antara lain halaman Beranda, halaman Publikasi, halaman Berita Resmi Statistik, halaman Tabel Statistik, halaman Senarai Rencana Terbit, halaman Berita Kegiatan, dan halaman Infografis.

Identifikasi Kebutuhan

Tahap selanjutnya yang perlu dilakukan adalah mengidentifikasi kebutuhan, yang didapatkan dari analisis SKPW 2021. Berdasarkan hasil analisis Gap dan IPA dalam SKPW 2021, kepuasan pengguna terhadap website BPS secara keseluruhan berada dalam kategori baik, namun masih ada beberapa atribut layanan yang kinerjanya belum memenuhi ekspektasi pengguna, antara lain kelengkapan metadata, kesesuaian produk yang ada pada website BPS dengan kebutuhan, kemudahan navigasi, kemudahan dalam memperoleh/mencari produk yang dibutuhkan, dan kehandalan fungsi pencarian (BPS, 2021). Selain itu, aspek pelayanan yang memiliki ekspektasi dan kinerja di bawah rata-rata adalah ketertarikan tampilan website, ketepatan

penyusunan tata letak, ketersediaan dua bahasa, dan kemudahan mendapatkan panduan penggunaan website (BPS, 2021).

Pengembangan Solusi Desain (Iterasi I)

Tahap selanjutnya dalam UCD adalah pengembangan solusi desain. Tahap ini dimulai dari pembuatan wireframe. Pembuatan wireframe ini dimulai dengan membuat sketsa di kertas. Menurut Gordon (2021) dari Nielsen Norman Group, terdapat 3 tahapan dalam membuat sketsa wireframe di kertas. Pertama adalah mengidentifikasi ukuran aspect ratio yang akan digunakan. Desain yang akan dibuat adalah desain untuk website BPS versi desktop sehingga ukuran yang digunakan adalah ukuran dengan aspect ratio 16:9. Selanjutnya adalah membuat gambar menu navigasi. Menu navigasi yang digunakan dalam perancangan desain BPS yang baru adalah menu navigasi di posisi atas website. Selanjutnya adalah mengidentifikasi dan membuat gambar elemen-elemen besar (Gordon, 2021). Misalnya pada halaman beranda, elemen besarnya adalah banner berupa gambar yang menjadi background section pencarian dan judul dari masing-masing section. Setelah membuat gambar elemen besar, selanjutnya adalah menambahkan detail (Gordon, 2021). Misalnya pada halaman beranda, detail yang perlu ditambahkan adalah dropdown menu dan button.

Sketsa yang telah dibuat di kertas tersebut kemudian dibuat di komputer dengan menggunakan tools diagrams.net. Wireframe halaman yang telah dibuat tersebut dipresentasikan secara rutin ke subject matter, kemudian subject matter memberikan komentar dan masukan terkait wireframe yang telah dibuat. Masukan dari subject matter dan keputusan hasil diskusi kemudian diakomodir dan diimplementasikan ke wireframe yang telah dibuat. Contoh rancangan wireframe dapat dilihat pada Gambar 4. Menu navigasi vertikal yang berada di sebelah kiri website BPS existing tidak ditampilkan di halaman beranda desain baru, hanya ditampilkan di halaman tabel statistik saja karena konten beranda cukup banyak, sementara navigasi

vertikal butuh lebih banyak ruang yang mengakibatkan ruang yang tersedia untuk konten beranda akan menjadi lebih sedikit (Laubheimer, 2021).

Pada awal perancangan wireframe, konten yang ditampilkan di halaman beranda dipilih dan diurutkan berdasarkan hasil SKPW 2021, namun kemudian ditambahkan juga sejumlah konten lain karena adanya beberapa pertimbangan khusus seperti pernyataan dari subject matter terkait permintaan deputi. Selanjutnya terkait halaman publikasi, pada tampilan daftar publikasi selain dalam bentuk list, pengguna juga diberi opsi untuk menampilkan dalam bentuk grid. Hal ini juga sesuai dengan poin ke-3 dari 10 Usability Heuristics Jakob Nielsen yaitu memberikan kontrol dan kebebasan pengguna (Nielsen, 2020).

Salah satu perubahan lain adalah penyajian tabel. Berdasarkan informasi dari subject matter, di website BPS existing terdapat 3 jenis tabel yaitu tabel statistik berdasarkan subjek, tabel dinamis, dan tabel SIMDASI. Ketiga jenis tabel ini berada di halaman yang terpisah. Subject matter menyatakan bahwa berhubung ke depannya BPS akan menggunakan klasifikasi kategori subjek yang baru, ketiga tabel tersebut akan diintegrasikan agar pengguna dapat langsung mengakses ketiganya di satu halaman saja. Terkait halaman tabel statistik, sempat didiskusikan bahwa tab metodologi dan konsep tabel seharusnya ada untuk masing-masing tabel dan tidak secara umum per subjek seperti yang ada di website BPS saat ini karena di satu subjek bisa saja terdapat indikator-indikator spesifik yang menggunakan metodologi dan konsep yang lebih khusus. Subject matter juga menginformasikan bahwa tab konsep dan metodologi yang saat ini ditampilkan di website existing sudah lama tidak di-update. Selain itu, dalam klasifikasi kategori subjek baru ada beberapa subjek yang memiliki penamaan dan konsep yang berbeda dengan yang digunakan saat ini. Setelah didiskusikan, keputusannya adalah memindahkan tab metodologi dan konsep ke halaman detail tabel statistik seperti pada Gambar 4.

Setelah pembuatan wireframe semua halaman yang masuk dalam cakupan telah selesai dan disetujui oleh subject matter, tahapan selanjutnya adalah pembuatan mockup. Berbeda dengan wireframe yang mengutamakan desain struktur, layout, dan tata letak konten, rancangan mockup sudah mencakup aspek desain visual seperti warna, gambar, dan tipografi (Adiseshiah, 2016). Untuk pembuatan mockup ini, tools yang digunakan adalah Figma. Sama seperti ketika desain wireframe, progress desain mockup juga rutin dipresentasikan ke subject matter yang kemudian memberikan komentar dan masukan. Dari segi warna, terdapat 2 warna utama yang digunakan di desain website BPS yang baru yaitu warna biru dan putih. Warna biru merepresentasikan kepercayaan, kompetensi, dan keandalan (Imtiaz, 2016). Warna putih melambangkan kemurnian, kebersihan, kesetaraan, kedamaian, dan kepolosan (Imtiaz, 2016). Selain itu ada juga warna kuning yang digunakan untuk banner pemberitahuan tepat di bawah navigasi utama (dapat dilihat di gambar 6). Warna ini merupakan warna komplementer dari warna biru. Warna komplementer utamanya digunakan untuk menarik perhatian ke suatu elemen (Imtiaz, 2016) sehingga cocok digunakan untuk banner pemberitahuan.

Ketika halaman beranda dari wireframe diimplementasikan ke dalam bentuk mockup, halamannya memiliki page scroll yang cukup panjang. Sebuah studi menunjukkan bahwa terdapat pola perhatian pengguna yang cenderung mengalami penurunan yang cukup tajam di tiap page fold berikutnya, sehingga semakin panjang page scroll yang dimiliki suatu halaman, semakin meningkat pula resiko kehilangan perhatian pengguna (Fessenden, 2018). Oleh karena itu, tampilan halaman beranda pada mockup dimodifikasi agar memiliki page scroll yang lebih sedikit dengan cara mengubah tampilan konten informasi terbaru (terdiri dari publikasi terbaru, BRS terbaru, berita terbaru, dan tabel statistik terbaru). Tampilan section ini pada wireframe terpisah menjadi section-section tersendiri, kemudian digabungkan

menjadi satu screen yang terbagi menjadi beberapa tab yang dapat dilihat pada Gambar 3. Tampilan mockup halaman beranda dapat dilihat pada gambar 4.

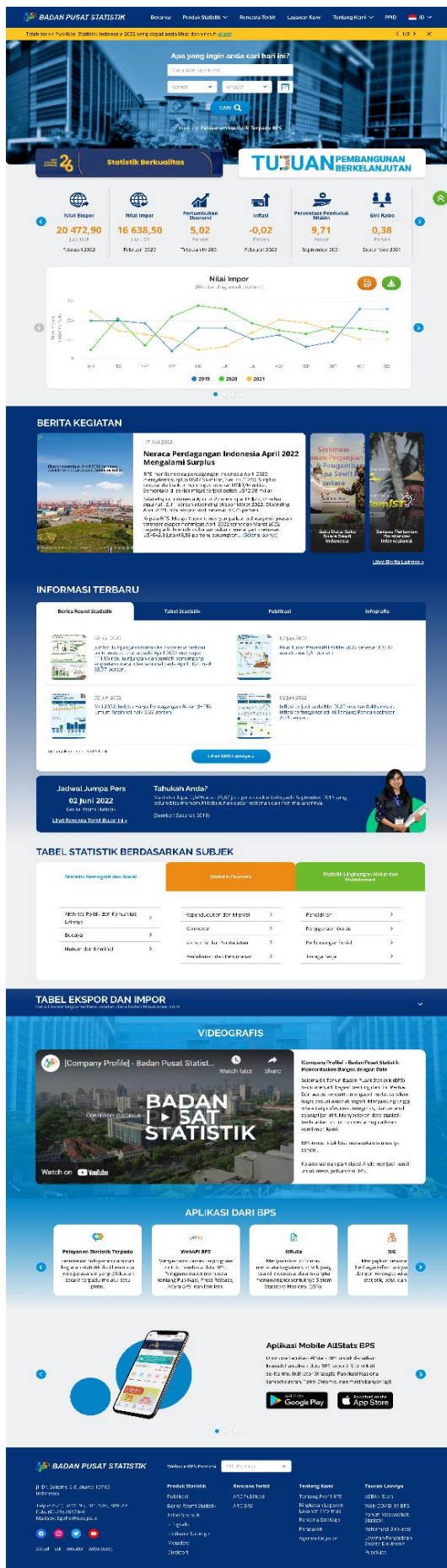
Tahap selanjutnya setelah mockup adalah pembuatan prototype. Dari rancangan mockup yang telah dibuat di figma, ditambahkan interaksi-interaksi serta animasi seperti efek hover pada button dan card, fungsi swap pada section indikator strategis dan grafik, navigasi ke halaman lain ketika sebuah menu di-klik, dan lain-lain. Untuk efek hover dilakukan dengan memanfaatkan fitur interactive component dari figma agar hanya perlu mendefinisikan efek yang diinginkan satu kali saja, kemudian dapat digunakan di seluruh elemen yang tersedia. Komponen-komponen interaktif yang telah dibuat setelahnya dirangkai menjadi satu halaman utuh yang kemudian dilakukan “wiring” untuk menghubungkan satu halaman ke halaman lain (Gambar 5).



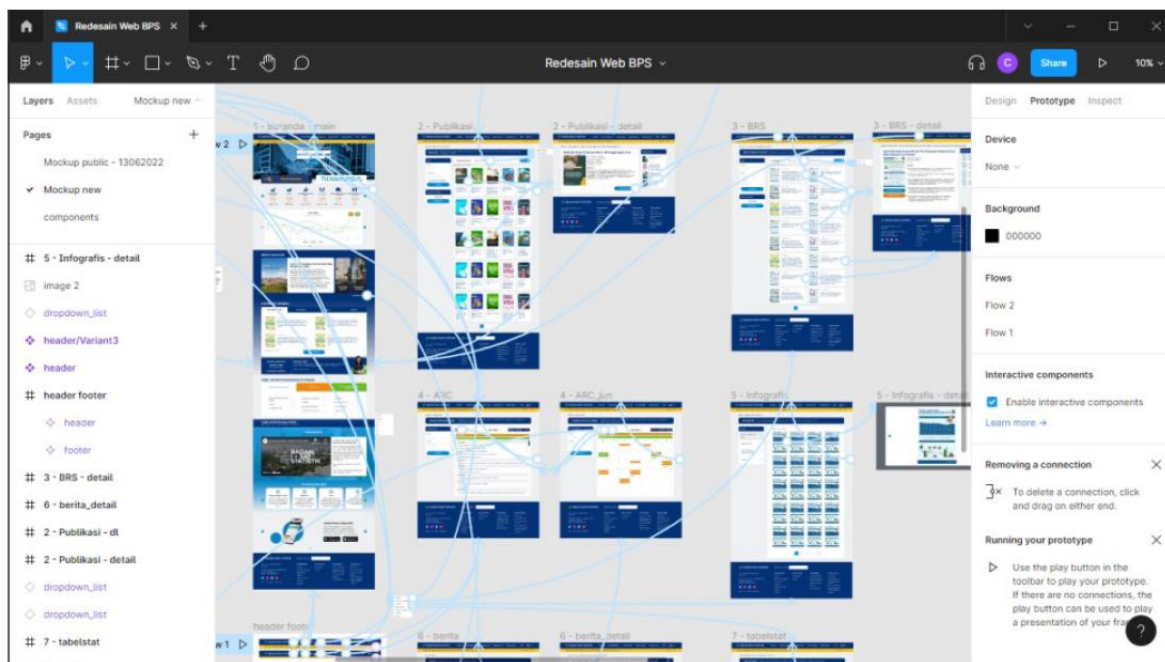
Gambar 2. Contoh Wireframe: Halaman Detail Tabel Statistik



Gambar 3. Perubahan Mockup Section Informasi Terbaru Halaman Beranda



Gambar 4. Contoh Mockup: Halaman Beranda



Gambar 5. Contoh *wiring* antarhalaman

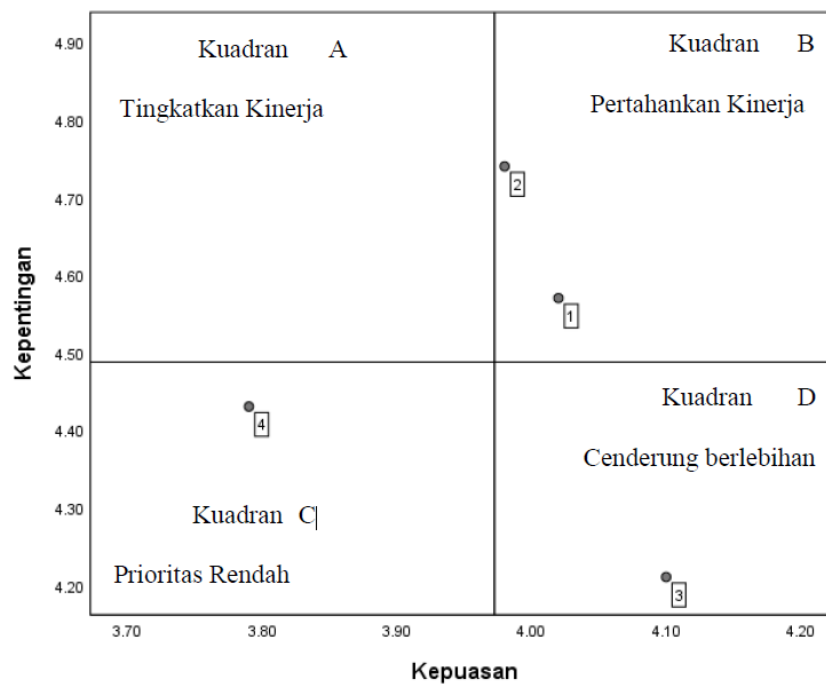
Evaluasi (Iterasi I)

Solusi desain yang sudah dibuat perlu dievaluasi untuk melihat apakah solusi sudah baik atau belum. Survei evaluasi dilakukan selama tiga hari pada tanggal 6 sampai 8 Juni 2022. Total responden yang mengisi kuesioner evaluasi adalah sebanyak 69 responden. Hasil analisis dengan IPA pada Gambar 6. Dapat dilihat dari Tabel 1 bahwa aspek kemudahan navigasi, kemudahan mencari produk, dan ketertarikan tampilan mengalami peningkatan jika dibandingkan dengan SKPW dengan rincian letak kuadran sebagai berikut:

- Aspek kemudahan navigasi dan kemudahan mencari produk yang pada SKPW 2021 berada di kuadran A (tingkatkan kinerja; perlu diprioritaskan perbaikannya) pada analisis IPA desain baru berada di kuadran B (pertahankan kinerja)
- Aspek ketertarikan tampilan yang pada SKPW 2021 berada di kuadran C (perlu perbaikan namun prioritas rendah) pada analisis IPA desain baru berada di kuadran D (cenderung berlebihan)
- Aspek ketepatan penyusunan tata letak pada analisis IPA desain baru masih

berada di kuadran C, sama seperti pada SKPW 2021.

Selain itu, dapat dilihat pada Gambar 7 bahwa hasil UEQ menunjukkan bahwa terdapat 1 skala yang berada pada kategori Good yaitu stimulasi serta 5 skala pada kategori Above Average yaitu daya tarik, kejelasan, efisiensi, ketepatan, dan kebaruan. Berdasarkan hasil ini, maka perancangan user interface masih perlu ditingkatkan lagi.

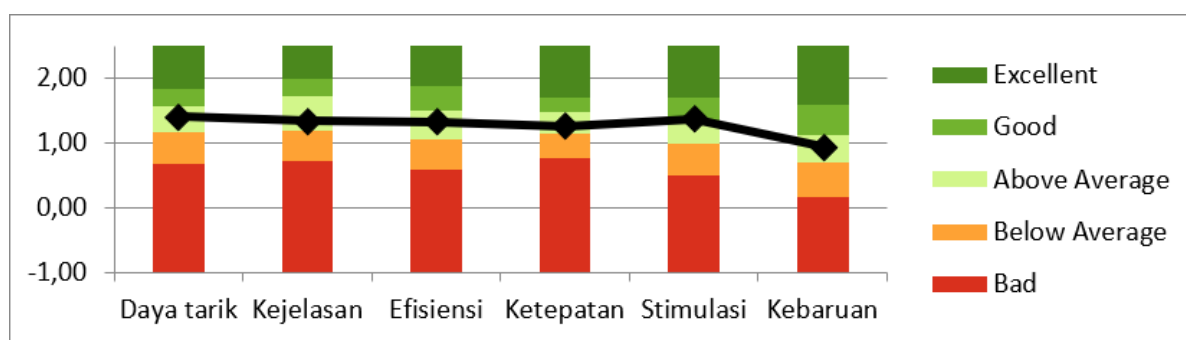


Gambar 6. Hasil analisis dengan IPA iterasi I (1) Kemudahan navigasi, (2) Kemudahan mencari produk yang dibutuhkan, (3) Ketertarikan tampilan, (4) Ketepatan penyusunan tata letak.

Tabel 1. Perubahan Nilai Tingkat Kepuasan dari SKPW dibandingkan Iterasi Pertama

Aspek	Tingkat Kepuasan di SKPW 2021	Tingkat Kepuasan di Iterasi I
Kemudahan Navigasi	3,91	4,02
Kemudahan mencari produk	3,91	4,98
Ketertarikan tampilan	3,90	4,10
Ketepatan penyusunan tata letak	3,87	3,79

Hijau = tingkat kepuasan meningkat; *Merah* = tingkat kepuasan menurun.



Gambar 7. Analisis UEQ Iterasi Pertama

Perubahan pada Solusi Desain (Iterasi II)

Sejumlah perubahan dilakukan pada rancangan solusi desain setelah mempertimbangkan masukan dari responden. Perubahan yang dilakukan

antara lain terletak pada urutan menu pada header, letak konten berita kegiatan, tampilan BPS App, perubahan warna latar, serta penambahan floating link untuk navigasi section beranda. Tampilan desain baru pada iterasi kedua dapat dilihat pada Gambar 8.



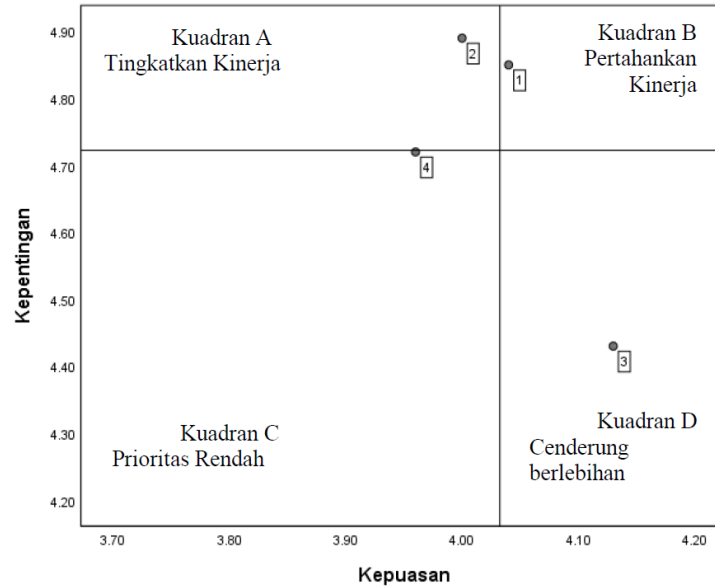
Gambar 8. Tampilan Halaman Beranda pada Iterasi Kedua

Evaluasi (Iterasi II)

Solusi desain pada iterasi ini juga perlu dievaluasi untuk melihat apakah solusi sudah baik atau belum. Survei evaluasi dilakukan selama tiga hari pada tanggal 14 sampai 16 Juni 2022. Total responden yang mengisi kuesioner evaluasi adalah sebanyak 76 responden. Hasil

analisis IPA iterasi II dapat dilihat pada Gambar 8 dengan rincian letak kuadran aspek-aspek pada analisis IPA yaitu:

- Aspek kemudahan navigasi dan kemudahan mencari produk yang pada iterasi pertama berada di kuadran B (pertahankan kinerja) pada analisis IPA iterasi kedua berada di kuadran A

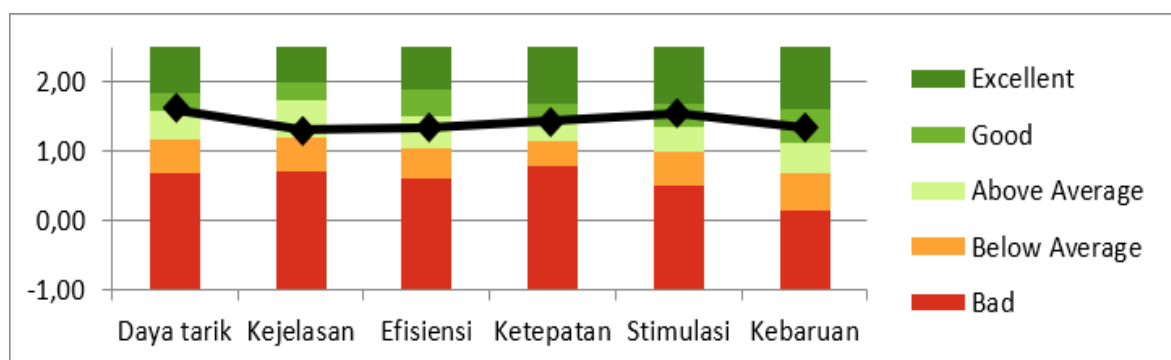


Gambar 8. Hasil analisis dengan IPA iterasi II (1) Kemudahan navigasi, (2) Kemudahan mencari produk yang dibutuhkan, (3) Ketertarikan tampilan, (4) Ketepatan penyusunan tata letak.

Tabel 2. Perubahan Nilai Tingkat Kepuasan dari SKPW hingga Iterasi Kedua

Aspek	Tingkat Kepuasan di SKPW 2021	Tingkat Kepuasan di Iterasi I	Tingkat Kepuasan di Iterasi II
Kemudahan Navigasi	3,91	4,02	4,04
Kemudahan mencari produk	3,91	4,98	4,00
Ketertarikan tampilan	3,90	4,10	4,13
Ketepatan penyusunan tata letak	3,87	3,79	3,96

Hijau = tingkat kepuasan meningkat; *Merah* = tingkat kepuasan menurun.



Gambar 9. Analisis UEO Iterasi Pertama

(tingkatkan kinerja; perlu diprioritaskan perbaikannya)

- Aspek ketertarikan tampilan yang pada iterasi pertama berada di kuadran D (cenderung berlebihan) pada analisis IPA iterasi kedua tetap berada di kuadran D
- Aspek ketepatan penyusunan tata letak pada analisis IPA iterasi kedua masih berada di kuadran C, sama seperti pada iterasi pertama

Dari hasil tersebut dapat disimpulkan bahwa aspek kemudahan navigasi, kemudahan mencari produk, dan ketepatan penyusunan tata letak masih dapat ditingkatkan lagi, namun jika dilihat dari nilai tingkat kepuasan, seluruh aspek telah mengalami peningkatan dibandingkan dengan hasil SKPW 2021 sebagaimana yang dapat dilihat pada tabel 1.

Dapat dilihat pada Gambar 9 bahwa hasil UEQ menunjukkan bahwa terdapat 3 skala yang berada pada kategori Good yaitu daya tarik, stimulasi, dan kebaruan, serta 3 skala pada kategori Above Average yaitu kejelasan, efisiensi, dan ketepatan. Skala daya tarik dan kebaruan yang pada iterasi sebelumnya berada di kategori Above Average, pada iterasi ini mengalami peningkatan menjadi kategori Good.


Setelah pengembangan solusi desain selesai, peneliti membuat pedoman UI untuk tampilan website BPS yang baru. Berikut adalah pedomannya.

Warna Website

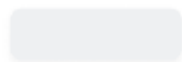


#EEF0F2	#043277	#0094DE
R 238 G 240 B 242	R 4 G 50 B 119	R 0 G 148 B 222

#EC8A14	#6BBB2C
R 236 G 138 B 20	R 107 G 187 B 44



Detail Warna Website



#EEF0F2
Background primer



#043277
Header, footer, sidebar header, tab info terbaru, ikon indikator strategis, warna button navigasi kalender ARC, teks judul section beranda, teks tautan



#0094DE
Button, tab tabel statistik



#EC8A14
Button, tab tabel statistik, legenda kalender ARC



#6BBB2C
Button, tab tabel statistik, legenda kalender ARC



#000000
Teks paragraf, teks judul



#FFFFFF
Background (BG) card, teks button, teks judul section beranda, BG side bar, BG dropdown menu, BG field



#A3A6AC
Teks placeholder

Tipografi



H1 header
Raleway Bold 32 pt



H2 header
Raleway Bold 24 pt

Default paragraph

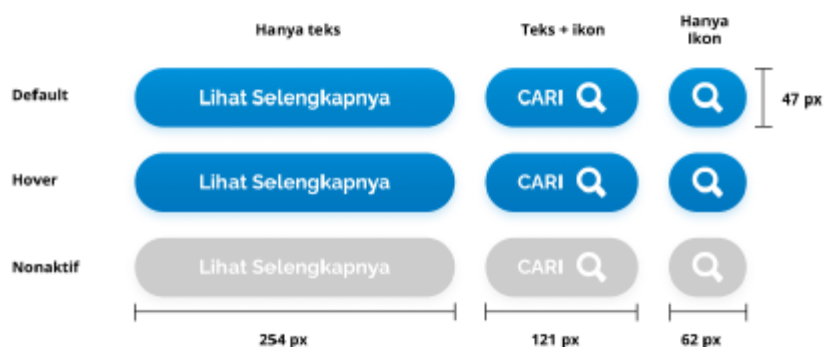
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus a enim sit amet sapien laoreet gravida sed iaculis velit. Mauris placerat nisi vel erat suscipit commodo. Nam ut mollis tortor, eu interdum leo. Nulla aliquet libero feugiat urna sagittis facilisis.

Open Sans Regular 16 pt

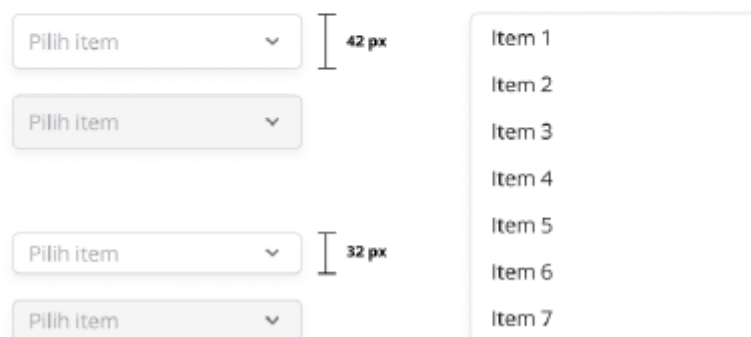
Detail Lorem ipsum dolor sit amet, consectetur adipiscing elit.

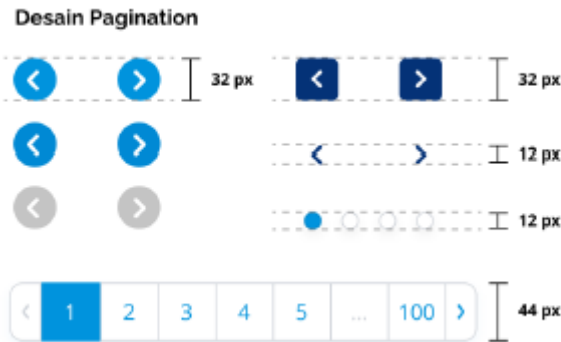
Open Sans Regular 12 pt

Desain Button



Desain Menu Dropdown





KESIMPULAN

Telah dilakukan perancangan kembali user interface beberapa halaman website BPS menggunakan metode User Centered Design dengan Survei Kepuasan Pengguna Website BPS 2021 sebagai dasar prioritas halaman yang dilakukan perancangan kembali, yaitu halaman beranda, publikasi, berita resmi statistik, senarai rencana terbit, tabel statistik, infografis, dan berita kegiatan. User interface baru yang telah dirancang telah dievaluasi ke pengguna sebanyak dua kali iterasi dengan iterasi terakhir memberikan hasil bahwa rancangan user interface website BPS yang baru dari segi tingkat kepuasan seluruh aspek pelayanannya yang dicakup dalam penelitian ini telah mengalami peningkatan dibandingkan dengan hasil SKPW 2021. Telah dihasilkan pula prototype beserta pedoman UI untuk tampilan website BPS yang baru.

DAFTAR PUSTAKA

- Abalo, J., Varela, J., & Manzano, V. (2007). Importance values for Importance–Performance Analysis: A formula for spreading out values derived from preference rankings. *Journal of Business Research*, 60(2), 115–121.
- Adishesiah. (2016). *Wireframes vs. Mockups: what's the best option?*. Retrieved from <https://www.justinmind.com/blog/wireframes-and-mockups-whats-the-best-option/>
- Ayhan, H. O. (2011). Non-probability Sampling Survey Methods. *International Encyclopedia of Statistical Science*, 14, 979–982.
- Badan Pusat Statistik. (2021). *Estimating Biomass and Biomass Change of Tropical Forest*. FAO Forestry Paper 134. FAO, Rome. Cited in <http://cdm.unfccc.int/Panels/ar>. [17 September 2007]
- Badan Pusat Statistik. (2021). *Hasil Survei Kepuasan Pengguna Website BPS 2021*. Jakarta: Badan Pusat Statistik.
- Badan Pusat Statistik. (2021). *Tentang Profil BPS*. Retrieved from <https://www.bps.go.id/menu/1/informasi-umum.html#masterMenuTab1>
- Budiu, R. & Moran, K. (2021). *How Many Participants for Quantitative Usability Studies: A Summary of Sample-Size Recommendations*. Retrieved from

- <https://www.nngroup.com/articles/summary-quant-sample-sizes/>
- Fessenden, T. (2018). *Scrolling and Attention*. Retrieved from <https://www.nngroup.com/articles/scrolling-and-attention/>
- Gordon, K. (2021) *How to Draw a Wireframe (Even if You Can't Draw)*. Retrieved from <https://www.nngroup.com/articles/draw-wireframe-even-if-you-cant-draw/>
- Imtiaz, S. (2016). *The Psychology Behind Web Design*. McMaster University. Canada.
- Interaction Design Foundation. (2016). *What is User Interface Design?*. Retrieved from <https://www.interaction-design.org/literature/topics/ui-design>
- Interaction Design Foundation. (2017). *What is User Centered Design?*. Retrieved from <https://www.interaction-design.org/literature/topics/user-centered-design>
- Jeff Johnson. (2008). *GUI Bloopers 2.0 - Common User Interface Design Don'ts and Dos*, Morgan Kaufmann, 2008
- Karnawan, G., Andryana, S., Komalasari, R. T. (2021). Implementasi User Experience Menggunakan Metode Design Thinking pada Prototype Aplikasi Cleanstic. *Jurnal TEKNOINFO*, 15(1), 61–66.
- Laubheimer, P. (2021). *Left-Side Vertical Navigation on Desktop: Scalable, Responsive, and Easy to Scan*. Retrieved from <https://www.nngroup.com/articles/vertical-nav/#:~:text=The%20main%20drawback%20to%20using,chrome%20ratio%20than%20horizontal%20navigation>
- Laugwitz, B., Schrepp, M., & Held, T. (2008). Construction and Evaluation of a User Experience Questionnaire. *Lecture Notes in Computer Science*, 5298, 63–76.
- Malewicz, M., & Malewicz, D. (2020). *Designing User Interfaces*. HYPE4. Washington.
- Martilla, J. A., & James, J. C. (1977). Importance-Performance Analysis. *Journal of Marketing*, 41(1), 77–79.
- Moran, K. (2019). *Usability Testing 101*. Retrieved from <https://www.nngroup.com/articles/usability-testing-101/>
- Nielsen, J. (2020). *10 Usability Heuristics for User Interface Design*. Retrieved from <https://www.nngroup.com/articles/ten-usability-heuristics/>
- Ormanovic, S., Ciric, A., Talovic, M., Alic, H., Jeleskovic, E., & Causevic, D. (2017). Importance-Performance Analysis: Different Approaches. *Acta Kinesiologica*, 11, 58–66.
- Sugiyono (2013). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta. Bandung.
- Wilbert O. Galitz. (2007). *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques (3rd Ed.)*, Wiley Publishing.

PERBANDINGAN ORDINAL FOREST DAN REGRESI LOGISTIK ORDINAL

Studi Kasus: Indeks Desa Membangun Provinsi Maluku Utara 2020

M. Yunus^{1,2}, Khairil Anwar Notodiputro¹, Bagus Sartono¹

¹IPB University, Bogor, Indonesia

²UIN Sulthan Thaha Saifuddin Jambi, Jambi, Indonesia

e-mail: muhammadyunus@apps.ipb.ac.id

Received: April 3, 2022; Revised: September 24, 2023; Accepted: October 6, 2023; Published: December 31, 2023

Abstrak

Terdapat banyak metode yang digunakan untuk mengidentifikasi prediktor penting dan memprediksi nilai dari peubah respons ordinal. Namun, terdapat metode prediksi untuk peubah respons ordinal yang tidak menggunakan sifat ordinal tersebut. Penelitian ini menggunakan metode ordinal forest dan metode regresi logistik ordinal. Nilai akurasi dan kappa metode ordinal forest pada setiap ulangan (5 ulangan) selalu lebih besar dari pada regresi logistik ordinal. Selanjutnya, nilai akurasi dan kappa setiap kelompok berdasarkan PDRB pada metode ordinal forest selalu lebih besar dari pada regresi logistik ordinal. Sehingga metode ordinal forest lebih baik digunakan pada data peringkat status indeks desa membangun Provinsi Maluku Utara 2020.

Kata kunci: indeks desa membangun, ordinal forest, peubah respons ordinal, regresi logistik ordinal

Abstract

There are many methods used to identify the important predictors and predict the value of the ordinal response variable. However, there are prediction methods for ordinal response variables that do not use the ordinal nature. This study uses the ordinal forest method and the ordinal logistic regression method. The accuracy and kappa value of the ordinal forest method in each replication (5 replications) is always greater than the ordinal logistic regression. Furthermore, the value of accuracy and kappa of each group based on GRDP in the ordinal forest method is always greater than the ordinal logistic regression. So that the ordinal forest method is better used in data on village development index in North Maluku Province 2020.

Keywords: ordinal forest, ordinal logistic regression, ordinal response variable, village development index

PENDAHULUAN

Terdapat banyak metode yang digunakan untuk mengidentifikasi prediktor penting dan memprediksi nilai dari peubah respons ordinal. Namun saat ini, terdapat metode prediksi untuk peubah respons ordinal yang tidak menggunakan sifat ordinal dari peubah respons tersebut. Peubah respons ordinal sering diperlakukan sebagai peubah multinomial ataupun nominal, menerapkan metode prediksi untuk peubah respons biner ke semua kombinasi pasangan kelas dari peubah respons ordinal. Salah satu metode statistika yang sering digunakan untuk memprediksi nilai dari peubah respons ordinal adalah regresi logistik ordinal (Atinafu et al., 2023; Baek et al., 2020; Moraes et al., 2019). Pendugaan parameter regresi logistik memakai metode kemungkinan maksimum (*maximum likelihood*) (Hosmer et al., 2013). Metode kemungkinan maksimum dapat diterapkan apabila antar amatan yang satu dengan amatan yang lain diasumsikan saling bebas. Pohon keputusan ordinal terbukti efektif untuk masalah klasifikasi di mana peubah kelas menunjukkan beberapa bentuk urutan ordinal (Singer et al., 2020).

Random forest klasik yang dikenalkan oleh (Breiman, 2001) belum menerapkan sifat ordinal pada peubah respons. Pada penelitian ini, diperkenalkan metode Ordinal Forest (OF). Ordinal forest merupakan metode prediksi inovatif untuk peubah respons ordinal yang berlaku untuk kovariat dimensi rendah dan dimensi tinggi. Metode OF menggunakan sifat ordinal dari peubah ordinal tersebut. Analog dengan kelas umum model regresi untuk peubah ordinal yang dikembangkan oleh (McCullagh, 1980) yang termasuk regresi probit terurut. Metode OF didasarkan pada gagasan tentang peubah respons kontinu yang mendasari peubah respon ordinal yang diamati (Hornung, 2019).

Ordinal forest cenderung mengungguli metode lain (*multi-class random forest, regression forest, dan naive ordinal forests*) dalam hal kinerja prediksi

dan terkait mengidentifikasi prediktor penting relatif sama baiknya dengan metode lain (Hornung, 2019). Namun, menurut (Nisa et al., 2021), metode terbaik adalah metode regresi logistik ordinal karena menghasilkan nilai akurasi yang lebih tinggi serta nilai sensitivitas dan spesifisitas yang lebih seimbang di setiap kategori IPK dibandingkan metode random forest ordinal. Oleh sebab itu, penelitian ini mengidentifikasi prediktor penting dan memprediksi nilai peubah respons ordinal dengan menggunakan metode ordinal forest dan sebagai pembanding digunakan juga metode regresi logistik ordinal.

Tujuan Pembangunan Berkelanjutan (TPB)/*Sustainable Development Goals* (SDGs) adalah pembangunan yang menjaga peningkatan kesejahteraan ekonomi, sosial, dan lingkungan masyarakat secara berkesinambungan (Kementerian PPN/Bappenas, 2020). Percepatan pencapaian tujuan SDGs secara nasional dapat didukung oleh formula SDGs Desa (Kurniawan et al., 2022). SDGs Desa dapat mewujudkan desa tanpa kemiskinan dan kelaparan, ekonomi tumbuh merata, peduli kesehatan, lingkungan, pendidikan, ramah perempuan, desa berjejaring, dan desa tanggap budaya (Permendes Nomor 13, 2020). Kementerian Desa, Pembangunan Daerah Tertinggal dan Transmigrasi Republik Indonesia (Kemendes PDTT RI) telah mengeluarkan Indeks Desa Membangun (IDM) untuk memberikan informasi dan status kemajuan desa guna mendukung pembangunan desa. IDM memiliki nilai pada interval (0,1) yang meliputi aspek sosial, ekonomi dan lingkungan (Astika & Subawa, 2021). Ilustrasi dalam penelitian ini menggunakan data peringkat status IDM Provinsi Maluku Utara tahun 2020.

METODOLOGI

1. Data

Pembangunan Daerah Tertinggal (PDT) merupakan suatu proses, upaya, dan tindakan secara terencana untuk meningkatkan kualitas masyarakat dan

wilayah yang merupakan bagian integral dari pembangunan nasional (Wahyuningrum, 2020). Pembangunan desa merupakan salah satu target penting dari kebijakan pemerintah di Indonesia (Utami & Wijayanto, 2022). Data penelitian yang dipakai merupakan data Peringkat Status Indeks Desa Membangun (IDM) Provinsi Maluku Utara Tahun 2020. Nilai Indeks Pembangunan Lingkungan di Provinsi Maluku Utara secara umum memenuhi klasifikasi rendah, sedang, dan tinggi. Nilai indeks tertinggi dimiliki oleh Kota Ternate (0,43) sedangkan nilai indeks terendah dimiliki oleh Kabupaten Halmahera Selatan (0,20) (Alfana et al., 2020). Data yang digunakan terdiri dari 1.199 observasi berupa desa/kelurahan dari 117 kecamatan di 10 kabupaten/kota pada Provinsi Maluku Utara. Terdapat 163 peubah di dalam data tersebut, peubah

Status IDM (Madjid et al., 2020) menjadi peubah respons dan sisanya merupakan peubah penjelas. Penelitian ini juga menggunakan data laju pertumbuhan Produk Domestik Regional Bruto sebagai dasar pengelompokan kategori laju pertumbuhan minus, rendah, sedang dan tinggi. Data diperoleh dari Badan Pusat Statistika (BPS) Provinsi Maluku Utara.

Menurut (Kemendes PDTT RI, 2020) klasifikasi/kategori Status Desa ditetapkan dengan ambang batas sebagai berikut:

1. Desa Sangat Tertinggal: $IDM \leq 0,4907$
2. Desa Tertinggal: $0,4907 < IDM \leq 0,5989$
3. Desa Berkembang: $0,5989 < IDM \leq 0,7072$
4. Desa Maju: $0,7072 < IDM \leq 0,8155$
5. Desa Mandiri: $IDM > 0,8155$

Tabel 1. Peubah Respons dan Peubah Penjelas

No.	Peubah Respons	Kategori	Skala
1.	Status Indeks Desa Membangun (IDM)	Sangat Tertinggal Tertinggal Berkembang Maju	Ordinal

No.	Peubah Penjelas	Kategori	Skala
1.	Jumlah Keluarga Pengguna Listrik		Rasio
2.	Penerangan di Jalan Utama Desa/Kelurahan	Ada, sebagian besar Ada, sebagian kecil Tidak ada	Ordinal
3.	Tempat Buang Sampah Sementara	Ada, digunakan Ada, tidak digunakan Tidak ada	Ordinal
4.	Keberadaan Bank Sampah	Ada Tidak ada	Nominal
5.	Keberadaan Sungai di Desa/Kelurahan	Ada Tidak ada	Nominal
6.	Jumlah Warga Penderita Gizi Buruk Tahun 2019		Rasio
7.	Keberadaan Angkutan Umum	Ada, dengan trayek tetap Ada, tanpa trayek tetap Tidak ada angkutan umum	Ordinal
⋮	⋮	⋮	⋮
160.	Jumlah Industri Barang dari Kayu		Rasio
161.	Jumlah Masjid		Rasio
162.	Kejadian/Bencana Alam (Banjir)	Ada Tidak ada	Nominal

Sumber: <https://idm.kemendes.go.id/view/detil/3/publikasi>

Tabel 1. Laju Pertumbuhan Produk Domestik Regional Bruto (PDRB) (Persen) Tahun 2020

No.	Wilayah	Laju Pertumbuhan	Kategori
1.	Halmahera Barat	0,62	Rendah
2.	Halmahera Tengah	26,34	Tinggi
3.	Kepulauan Sula	0,41	Rendah
4.	Halmahera Selatan	16,22	Tinggi
5.	Halmahera Utara	-0,15	Minus
6.	Halmahera Timur	0,57	Rendah
7.	Pulau Morotai	2,68	Sedang
8.	Pulau Taliabu	1,06	Sedang
9.	Ternate	-0,85	Minus
10.	Tidore Kepulauan	1,99	Sedang
	Maluku Utara	4,92	

Sumber: <https://malut.bps.go.id>

Tabel 1. menunjukkan peubah respons dan peubah penjelas yang digunakan dalam penelitian ini. Tabel 2. menyajikan laju Pertumbuhan Produk Domestik Regional Bruto (PDRB) Menurut kabupaten/kota Provinsi Maluku Utara Atas Dasar Harga Konstan (Persen) Tahun 2020.

Metode Penelitian

Tahapan-tahapan analisis data menggunakan perangkat lunak R dengan paket ordinalForest, VGAM, dan caret sebagai berikut:

1. Eksplorasi Data
2. Analisis Ordinal Forest (OF)
Analisis dengan menggunakan ordinal forest, algoritma OF diimplementasikan pada paket R ordinalForest yang tersedia di CRAN dalam versi 2.4-2 (Hornung, 2021). Keakuratan tes diagnostik sering dinilai oleh dua probabilitas bersyarat yaitu sensitivitas dan spesifisitas (Agresti, 2019; Wulandari, 2015). Penilaian diagnostik Y memiliki skala ordinal, sebagai berikut:
spesifisitas = $P(Y \leq j | x=0)$,
sensitivitas = $P(Y > j | x=1)$
3. Analisis Regresi Logistik Ordinal
Persamaan (1) merupakan persamaan analisis dengan menggunakan regresi logistik ordinal (Agresti, 2010) diberikan:

$$P(Y \leq j | \mathbf{x}) = \pi_1(\mathbf{x}) + \dots + \pi_j(\mathbf{x}),$$

$$j = 1, \dots, J.$$

$$\text{logit} [P(Y \leq j | \mathbf{x})] = \log \frac{P(Y \leq j | \mathbf{x})}{1 - P(Y \leq j | \mathbf{x})}$$

$$\text{logit} [P(Y \leq j | \mathbf{x})]$$

$$= \log \frac{\pi_1(\mathbf{x}) + \dots + \pi_j(\mathbf{x})}{\pi_{j+1}(\mathbf{x}) + \dots + \pi_J(\mathbf{x})} \quad (1)$$

$$j = 1, \dots, J - 1$$

selanjutnya, juga dilakukan tes diagnostik yaitu sensitivitas dan spesifisitas.

4. Perbandingan Analisis Ordinal Forest dan Regresi Logistik Ordinal
Perbandingan Analisis Ordinal Forest dan Regresi Logistik Ordinal menggunakan nilai akurasi dan kappa. Berikut ini perhitungan nilai kappa (Cohen, 1968) pada persamaan (2), diberikan:

$$K_w = \frac{\sum_{i=1}^J \sum_{j=1}^J w_{ij} p_{oij} - \sum_{i=1}^J \sum_{j=1}^J w_{ij} p_{cij}}{1 - \sum_{i=1}^J \sum_{j=1}^J w_{ij} p_{cij}} \quad (2)$$

5. Ukuran Kepentingan Peubah
Ukuran yang cocok digunakan dalam (Janitza et al., 2016) untuk menilai peringkat peubah dalam hal ini diperoleh dengan VIM (Variabel Importance Measures) pada persamaan (3). Ukuran

kepentingan peubah untuk kovariat j diberikan sebagai berikut:

$$VIM_j^M = \frac{1}{ntree} \sum_{t=1}^{ntree} (MP_{tj} - M_{tj}), \quad (3)$$

$ntree$: Jumlah pohon di hutan,

M_{tj} : Kesalahan pohon t sebelum mengubah nilai variabel prediktor X_j ketika memprediksi observasi.

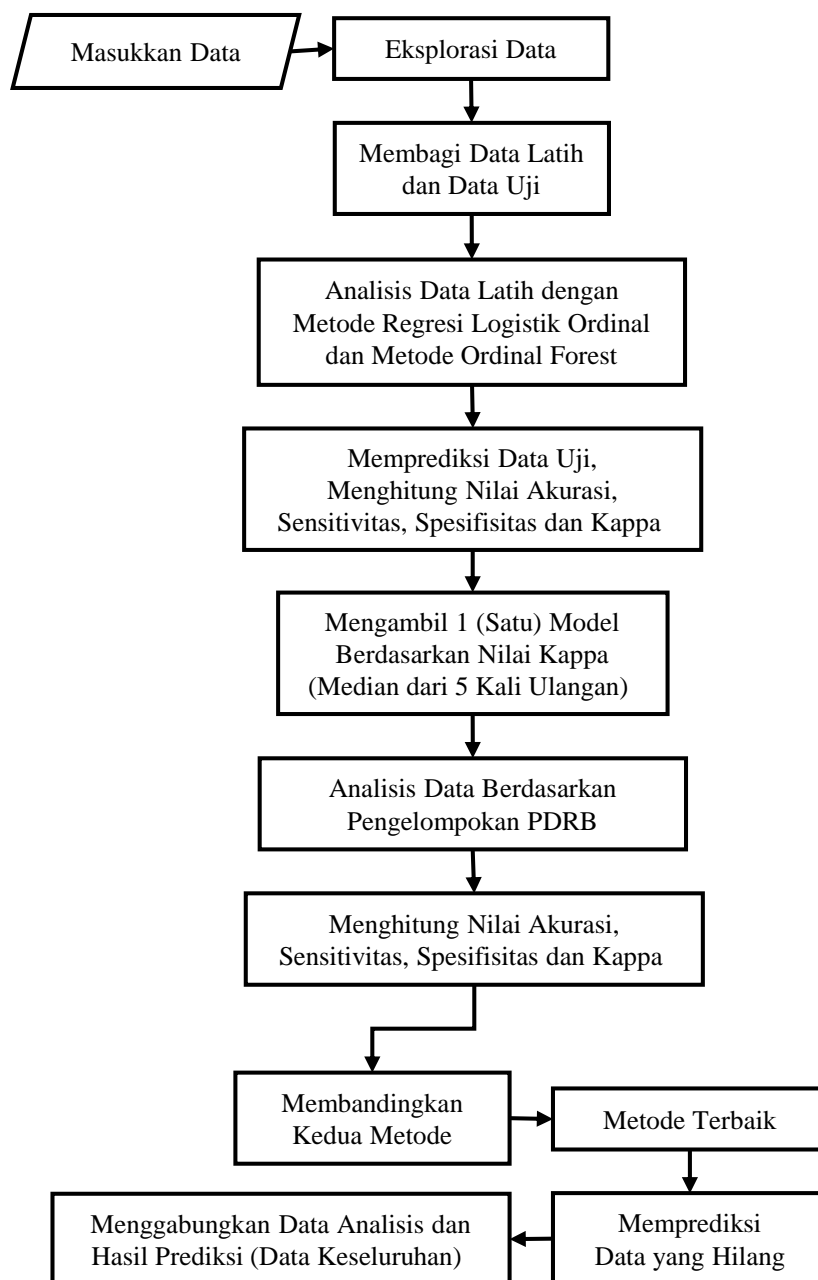
MP_{tj} : Kesalahan pohon t sesudah mengubah nilai variabel prediktor X_j ketika memprediksi observasi.

6. Prediksi Status IDM yang Mengalami Missing Value (Nilai yang Hilang)

Prediksi terhadap status IDM yang mengalami Missing Value (Nilai yang Hilang) pada data Peringkat Status Indeks Desa Membangun (IDM) Provinsi Maluku Utara Tahun 2020.

2. Prosedur Analisis Data

Prosedur analisis data dalam penelitian ini, disajikan pada Gambar 1. Diagram Alir Prosedur Analisis Data, sebagai berikut:



Gambar 1. Diagram Alir Prosedur Analisis Data

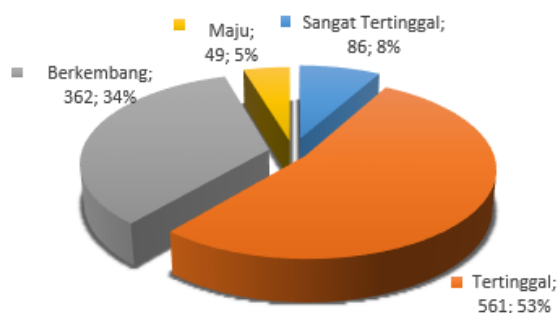
Tabel 3. Kategori Sebenarnya Dan Hasil Prediksi Analisis OF

Prediksi	Kategori Sebenarnya			
	Sangat Tertinggal	Tertinggal	Berkembang	Maju
Sangat Tertinggal	4	1	0	0
Tertinggal	12	96	24	0
Berkembang	0	14	50	3
Maju	0	0	1	7
Total	16	111	75	10

HASIL DAN PEMBAHASAN

1. Eksplorasi Data

Data peringkat status Indeks Desa Membangun (IDM) Provinsi Maluku Utara Tahun 2020. Data terdiri dari 1.199 observasi (desa/kelurahan), namun terdapat 141 desa/kelurahan mengalami *missing value* (nilai yang hilang) terkait dengan peubah respons Status IDM, sehingga tidak diikutsertakan ke dalam analisis. Sisa desa/kelurahan yang digunakan dalam penelitian ini adalah 1.058 desa/kelurahan. Selanjutnya, nilai yang hilang dari peubah Status IDM akan diprediksi menggunakan metode terbaik yang diperoleh pada penelitian ini. Gambar 2 menunjukkan bahwa sebagian besar status IDM di Provinsi Maluku Utara tahun 2020 dikategorikan sebagai desa tertinggal yaitu sebesar 53% (561 observasi).



Gambar 2. Sebaran Status IDM di 1.058 Desa/Kelurahan

2. Analisis Ordinal Forest

Analisis ordinal forest (OF) menggunakan data latih sebesar 80% (846 desa/kelurahan) dan data uji sebesar 20% (212 desa/kelurahan). Proses analisis tersebut dilakukan sebanyak 5 kali ulangan

dengan data latih dan data uji yang berbeda dan dengan persentase yang sama. Diperoleh nilai kappa dari setiap ulangan dan diambil median dari 5 nilai kappa tersebut. Tabel 3. menunjukkan kategori sebenarnya dan hasil prediksi menggunakan analisis ordinal forest (OF) dengan nilai kappa sebesar 0,5356.

Berdasarkan Tabel 3. menunjukkan bahwa sebanyak 157 desa/kelurahan diprediksi benar oleh metode tersebut. Terdapat 55 desa/kelurahan yang diprediksi tidak sesuai dengan kategori sebenarnya. Selanjutnya, dari 55 desa/kelurahan tersebut, terdapat 27 desa/kelurahan yang diprediksi lebih tinggi dari kategori yang sebenarnya dan 28 desa/kelurahan yang diprediksi lebih rendah dari kategori sebenarnya. Tabel 4. menyajikan nilai evaluasi metode ordinal forest berupa nilai akurasi, nilai sensitivitas dan nilai spesifisitas setiap kategori status IDM.

Berdasarkan Tabel 4. dapat dilihat bahwa nilai akurasi menggunakan metode logistik ordinal sebesar 74,06% artinya sebanyak 157 desa/kelurahan dari 212 desa/kelurahan diprediksi sesuai dengan kategorinya, sisanya 55 desa/kelurahan diprediksi tidak sesuai dengan kategorinya. Nilai sensitivitas untuk kategori sangat tertinggal sebesar 25,00% artinya sebanyak 4 desa/kelurahan dari 16 desa/kelurahan diprediksi sesuai dengan kategorinya, sisanya 12 desa/kelurahan diprediksi tidak sesuai dengan kategorinya dan nilai spesifisitas sebesar 99,49%. Nilai sensitivitas untuk kategori tertinggal sebesar 86,49% artinya sebanyak 96 desa/kelurahan dari 111 desa/kelurahan diprediksi sesuai dengan kategorinya, sisanya 15 desa/kelurahan diprediksi tidak

Tabel 4. Nilai Evaluasi Metode Ordinal Forest (%)

Nilai	Kategori Status IDM	Status IDM
Akurasi		74,06
Sensitivitas	Sangat Tertinggal	25,00
	Tertinggal	86,49
	Berkembang	66,67
	Maju	70,00
Spesifisitas	Sangat Tertinggal	99,49
	Tertinggal	64,36
	Berkembang	87,59
	Maju	99,51

sesuai dengan kategorinya dan nilai spesifisitas sebesar 64,36%. Nilai sensitivitas untuk kategori berkembang sebesar 66,67% artinya sebanyak 50 desa/kelurahan dari 75 desa/kelurahan diprediksi sesuai dengan kategorinya, sisanya 25 desa/kelurahan diprediksi tidak sesuai dengan kategorinya dan nilai spesifisitas sebesar 87,59%. Nilai sensitivitas untuk kategori maju sebesar 70,00% artinya sebanyak 7 desa/kelurahan dari 10 desa/kelurahan diprediksi sesuai dengan kategorinya, sisanya 3 desa/kelurahan diprediksi tidak sesuai dengan kategorinya dan nilai spesifisitas sebesar 99,51%.

3. Analisis Regresi Logistik Ordinal

Analisis regresi logistik ordinal menggunakan data latih sebesar 80% (846 desa/kelurahan) dan data uji sebesar 20% (212 desa/kelurahan). Proses analisis tersebut dilakukan sebanyak 5 kali ulangan dengan data latih dan data uji yang berbeda dan dengan persentase yang sama. Diperoleh nilai kappa dari setiap ulangan dan diambil median dari 5 nilai kappa tersebut. Tabel 5. menunjukkan kategori sebenarnya dan hasil prediksi

menggunakan analisis regresi logistik ordinal dengan nilai kappa sebesar 0,3922.

Berdasarkan Tabel 5. menunjukkan bahwa sebanyak 137 desa/kelurahan diprediksi benar oleh metode tersebut. Terdapat 75 desa/kelurahan yang diprediksi tidak sesuai dengan kategori sebenarnya. Selanjutnya, dari 75 desa/kelurahan tersebut, terdapat 37 desa/kelurahan yang diprediksi lebih tinggi dari kategori yang sebenarnya dan 38 desa/kelurahan yang diprediksi lebih rendah dari kategori sebenarnya. Tabel 6. menyajikan nilai evaluasi metode regresi logistik ordinal berupa nilai akurasi, nilai sensitivitas dan nilai spesifisitas setiap kategori status IDM.

Berdasarkan Tabel 6. dapat dilihat bahwa nilai akurasi menggunakan metode logistik ordinal sebesar 64,62% artinya sebanyak 137 desa/kelurahan dari 212 desa/kelurahan diprediksi sesuai dengan kategorinya, sisanya 75 desa/kelurahan diprediksi tidak sesuai dengan kategorinya. Nilai sensitivitas untuk kategori sangat tertinggal sebesar 31,25% artinya sebanyak

Tabel 5. Kategori Sebenarnya Dan Hasil Prediksi Analisis Regresi Logistik Ordinal

Prediksi	Kategori Sebenarnya			
	Sangat Tertinggal	Tertinggal	Berkembang	Maju
Sangat Tertinggal	5	5	2	0
Tertinggal	11	85	29	0
Berkembang	0	20	39	2
Maju	0	1	5	8
Total	16	111	75	10

Tabel 6. Nilai Evaluasi Metode Regresi Logistik Ordinal (%)

Nilai	Kategori Status IDM	Status IDM
Akurasi		64,62
Sensitivitas	Sangat Tertinggal	31,25
	Tertinggal	76,58
	Berkembang	52,00
	Maju	80,00
Spesifisitas	Sangat Tertinggal	96,43
	Tertinggal	60,40
	Berkembang	83,94
	Maju	97,03

5 desa/kelurahan dari 16 desa/kelurahan diprediksi sesuai dengan kategorinya, sisanya 11 desa/kelurahan diprediksi tidak sesuai dengan kategorinya dan nilai spesifisitas sebesar 96,43%. Nilai sensitivitas untuk kategori tertinggal sebesar 76,58% artinya sebanyak 85 desa/kelurahan dari 111 desa/kelurahan diprediksi sesuai dengan kategorinya, sisanya 26 desa/kelurahan diprediksi tidak sesuai dengan kategorinya dan nilai spesifisitas sebesar 60,40%. Nilai sensitivitas untuk kategori berkembang sebesar 52,00% artinya sebanyak 39 desa/kelurahan dari 75 desa/kelurahan diprediksi sesuai dengan kategorinya, sisanya 36 desa/kelurahan diprediksi tidak sesuai dengan kategorinya dan nilai spesifisitas sebesar 83,94%. Nilai sensitivitas untuk kategori maju sebesar 80,00% artinya sebanyak 8 desa/kelurahan dari 10 desa/kelurahan diprediksi sesuai dengan kategorinya, sisanya 2 desa/kelurahan diprediksi tidak sesuai dengan kategorinya dan nilai spesifisitas sebesar 97,03%.

4. Perbandingan Ordinal Forest dan Regresi Logistik Ordinal

Pemodelan terbaik dipilih berdasarkan nilai akurasi dan kappa.

Metode yang mempunyai nilai akurasi dan kappa terbesar merupakan metode yang terbaik. Tabel 7. menyajikan perbandingan ordinal forest dan regresi logistik ordinal. Berdasarkan Tabel 7. bahwa nilai akurasi dan kappa setiap ulangan dari metode ordinal forest lebih besar dari pada regresi logistik ordinal. Didapatkan metode ordinal forest lebih baik digunakan pada data peringkat status indeks desa membangun Provinsi Maluku Utara tahun 2020. Model yang diambil adalah 1 (satu) model dengan set.seed (137) berdasarkan nilai kappa (median dari 5 kali ulangan). Selanjutnya, dilakukan analisis data Status Indeks Desa Membangun Provinsi Maluku Utara Tahun 2020 yang telah dikelompokkan berdasarkan laju pertumbuhan PDRB. Tabel 8. menyajikan perbandingan regresi logistik ordinal dan ordinal forest pada data pengelompokan berdasarkan PDRB tahun 2020.

Berdasarkan Tabel 8. bahwa nilai akurasi dan kappa setiap kelompok pada metode ordinal forest selalu lebih besar dari pada regresi logistik ordinal. Sehingga didapatkan metode ordinal forest lebih baik digunakan pada data pengelompokan berdasarkan PDRB di status indeks desa membangun Provinsi Maluku Utara tahun 2020.

Tabel 7. Perbandingan Ordinal Forest (OF) dan Regresi Logistik Ordinal (RLO)

Metode	Nilai	set.seed (17)	set.seed (27)	set.seed (67)	set.seed (137)	set.seed (257)	Media
OF	Akurasi	0,7500	0,7547	0,7028	0,7406	0,7311	0,7406
	Kappa	0,5415	0,5414	0,4565	0,5356	0,5128	0,5356
RLO	Akurasi	0,6981	0,6557	0,6462	0,6462	0,6934	0,6557
	Kappa	0,4642	0,3833	0,3695	0,3922	0,4614	0,3922

Tabel 8. Perbandingan Ordinal Forest (OF) dan Regresi Logistik Ordinal (RLO) pada Data Pengelompokan Berdasarkan PDRB Tahun 2020

Metode	Nilai	Laju Pertumbuhan PDRB			
		Minus	Rendah	Sedang	Tinggi
OF	Akurasi	0,9490	0,9568	0,9569	0,9314
	Kappa	0,9143	0,9252	0,9233	0,8621
RLO	Akurasi	0,7398	0,7666	0,7703	0,7582
	Kappa	0,5397	0,5856	0,5863	0,4849

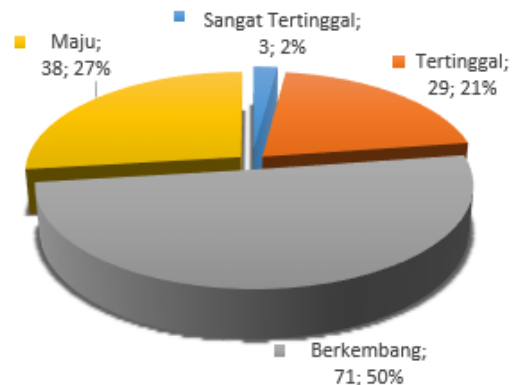
5. Ukuran Kepentingan Peubah

Ukuran kepentingan peubah penjelas yang digunakan pada penelitian ini adalah VIM (*Variable Importance Measures*). Tingkat kepentingan 7 peubah penjelas numerik yang lebih tinggi dibandingkan dengan peubah penjelas lainnya adalah jarak restoran/rumah makan, jarak warung/kedai makanan minuman, jarak tempat praktik dokter, jarak kantor bupati/walikota, jarak minimarket/swalayan, jumlah keluarga pengguna listrik dan jarak apotek. Selanjutnya, tingkat kepentingan 7 peubah penjelas kategorik yang lebih tinggi dibandingkan dengan peubah penjelas lainnya adalah tempat buang sampah keluarga, sumber air untuk mandi/cuci, lalu lintas (transport) dari/ke desa/kelurahan, sumber air untuk minum, keberadaan angkutan umum, bahan bakar untuk memasak dan sinyal telepon seluler/handphone. Hal ini sesuai dengan penelitian yang dilakukan (Prasetya et al., 2019) bahwa nilai Indeks Jaminan Ekonomi dan Sosial cenderung tinggi.

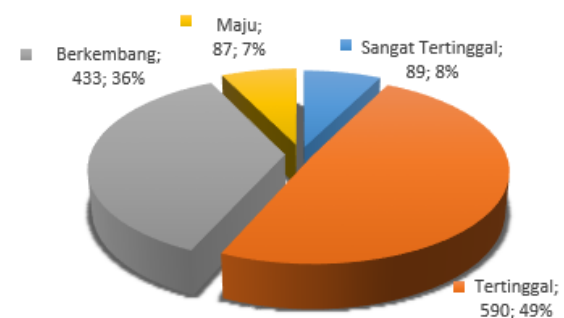
6. Prediksi Status IDM yang Mengalami Missing Value (Nilai yang Hilang)

Terdapat 141 desa/kelurahan mengalami *missing value* (nilai yang hilang) terkait dengan peubah respons Status IDM, sehingga dilakukan prediksi dengan metode terbaik yaitu ordinal forest. Gambar 3. menunjukkan sebaran hasil prediksi status IDM di 141 desa/kelurahan. Sebagian besar desa/kelurahan yang mengalami nilai yang hilang diprediksi sebagai desa yang berstatus berkembang yaitu sebesar 71 desa/kelurahan (50%). Hanya ada 3 desa/kelurahan (2%) yang

diprediksi sebagai desa yang berstatus sangat tertinggal. Gambar 4. menunjukkan sebaran status IDM di 1.199 desa/kelurahan pada Provinsi Maluku Utara tahun 2020 (secara keseluruhan termasuk hasil prediksi).



Gambar 3. Sebaran Status IDM di 141 Desa/Kelurahan



Gambar 4. Sebaran Status IDM di 1.199 Desa/Kelurahan

SIMPULAN

Nilai akurasi dan kappa metode ordinal forest pada setiap ulangan (5 ulangan) selalu lebih besar dari pada regresi logistik ordinal. Selanjutnya, nilai akurasi dan kappa setiap kelompok berdasarkan PDRB pada metode ordinal forest selalu

lebih besar dari pada regresi logistik ordinal. Sehingga disimpulkan bahwa metode ordinal forest lebih baik digunakan pada data peringkat status indeks desa membangun Provinsi Maluku Utara tahun 2020. Tingkat kepentingan peubah penjelas numerik dan kategorik yang tertinggi dibandingkan dengan peubah penjelas lainnya adalah jarak restoran/rumah makan dan tempat buang sampah keluarga. Hasil prediksi nilai yang hilang sebanyak 141 desa/kelurahan dengan metode ordinal forest, sebagian besar desa/kelurahan yang mengalami nilai yang hilang diprediksi sebagai desa yang berstatus berkembang yaitu sebesar 71 desa/kelurahan (50%). Hanya ada 3 desa/kelurahan (2%) yang diprediksi sebagai desa yang berstatus sangat tertinggal.

DAFTAR PUSTAKA

- Agresti, Alan. (2010). *Analysis of Ordinal Categorical Data* (Second Edition). John Wiley and Sons.
- Agresti, Alan. (2019). An Introduction to Categorical Data Analysis. In *Wiley Series in Probability and Statistics* (Third Edition, Vol. 170, Issue 4). https://doi.org/10.1111/j.1467-985x.2007.00506_2.x
- Alfana, A. M. F., Kusmanto, Y., & Pitoyo, A. J. (2020). *Analisis Indeks Pembangunan Lingkungan di Provinsi Maluku Utara Analisis Sumberdaya Manusia View project do an exercise View project*. <https://www.researchgate.net/publication/344778695>
- Astika, A. N., & Subawa, N. S. (2021). Evaluasi Pembangunan Desa Berdasarkan Indeks Desa Membangun. *Jurnal Ilmiah Muqoddimah : Jurnal Ilmu Sosial, Politik Dan Humaniora*, 5(2), 223–232. <http://jurnal.um-tapsel.ac.id/index.php/muqoddimah>
- Atinafu, B. T., Tarekegn, F. N., & Kebede, W. M. (2023). Assessment of the Level of Social Support and Associated Factors Among Cancer Patients in the Comprehensive Cancer Center at Ethiopia: Ordinal Logistic Regression Analysis Level of Social Support and Associated Factors Among Cancer Patients. *Heliyon*, 9(5). <https://doi.org/10.1016/j.heliyon.2023.e15688>
- Baek, W., Jang, Y., Park, C. G., & Moon, M. (2020). Factors Influencing Satisfaction with Patient-Controlled Analgesia Among Postoperative Patients Using a Generalized Ordinal Logistic Regression Model. *Asian Nursing Research*, 14(2), 73–81. <https://doi.org/10.1016/j.anr.2020.03.001>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.
- Cohen, J. (1968). Weighed Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit. *Psychological Bulletin*, 70(4), 213–220.
- Hornung, R. (2019). Ordinal Forests. *Journal of Classification*, 37(1), 4–17. <https://doi.org/10.1007/s00357-018-9302-x>
- Hornung, R. (2021). Ordinal Forests: Prediction and Variable Ranking with Ordinal Target Variables. In *R package version 2.4-2*. Springer. <https://doi.org/10.1007/s00357-018-9302-x>
- Hosmer, D. W., Lemeshow, Stanley., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (Third Edition). New Jersey (US): John Wiley and Sons.
- Janitza, S., Tutz, G., & Boulesteix, A. L. (2016). Random Forest for Ordinal Responses: Prediction and Variable Selection. *Computational Statistics and Data Analysis*, 96, 57–73. <https://doi.org/10.1016/j.csda.2015.10.005>
- Kemendes PDTT RI. (2020). *Standar Operasional Prosedur (SOP) Update Data Indeks Desa Membangun Tahun 2020*. Direktorat Jenderal Pembangunan dan Pemberdayaan Masyarakat Desa: Kementerian Desa, Pembangunan Daerah Tertinggal dan Transmigrasi, Republik Indonesia.
- Kementerian PPN/Bappenas. (2020). *Pedoman Teknis Penyusunan Rencana Aksi Tujuan Pembangunan Berkelanjutan (TPB)/ Sustainable Development Goals (SDGs)*. Kementerian Perencanaan Pembangunan

- Nasional / Badan Perencanaan dan Pembangunan Nasional (BAPPENAS).
- Kurniawan, E., Amidi, Gunawan, Susilowati, N., Paranti, L., & Santi W, D. G. (2022). *Panduan UNNES GIAT Penguatan Generasi Milenial Mendukung SDGs Desa*.
- Madjid, T., Rachmawaty, R., Fachri, M., Wikantosa, B., Uguy, N. S., Muklis, Agusta, I., Hariyanti, Mustakim, Dwiatmojo, R., Azizah, K., & Sukoyo. (2020). *Peringkat Status Indeks Desa Membangun (IDM)* (2020th ed.). Direktorat Jenderal Pembangunan dan Pemberdayaan Masyarakat Desa: Kementerian Desa, Pembangunan Daerah Tertinggal dan Transmigrasi, Republik Indonesia.
- McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(2), 109–142.
- Moraes, J. O., Cruz, E. A., Pinheiro, Í., Oliveira, T. C. M., Alvarenga, V., Sant'Ana, A. S., & Magnani, M. (2019). An Ordinal Logistic Regression Approach to Predict the Variability on Biofilm Formation Stages by Five Salmonella Enterica Strains on Polypropylene and Glass Surfaces as Affected by pH, Temperature and NaCl. *Food Microbiology*, 83, 95–103. <https://doi.org/10.1016/j.fm.2019.04.012>
- Nisa, Z. I., Soleh, A. M., & Wijayanto, H. (2021). Identifikasi Faktor-Faktor yang Memengaruhi Prestasi Mahasiswa Menggunakan Regresi Logistik Ordinal dan Random Forest Ordinal. *Xplore: Journal of Statistics*, 10(1), 88–101. <https://doi.org/10.29244/xplore.v10i1.465>
- Permendesa Nomor 13. (2020). *Peraturan Menteri Desa, Pembangunan Daerah Tertinggal dan Transmigrasi Republik Indonesia*.
- Prasetya, Y. E., Hidayat, A. R. T., & Dinanti, D. (2019). Village Development Index of Probolinggo Coastal Villages Case study: Bhinor Village, Paiton District. *IOP Conference Series: Earth and Environmental Science*, 328(1), 1–8. <https://doi.org/10.1088/1755-1315/328/1/012056>
- Singer, G., Anuar, R., & Ben-Gal, I. (2020). A Weighted Information-Gain Measure for Ordinal Classification Trees. *Expert Systems with Applications*, 152. <https://doi.org/10.1016/j.eswa.2020.113375>
- Utami, N. A., & Wijayanto, A. W. (2022). Classification of Village Development Index at Regency/Municipality Level Using Bayesian Network Approach with K-Means Discretization. *Jurnal Aplikasi Statistika & Komputasi Statistik, Khusus*, 95–206.
- Wahyuningrum, V. (2020). Penerapan Radial Basis Function Neural Network dalam Pengklasifikasian Daerah Tertinggal di Indonesia. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 12(1), 37–54.
- Wulandari, I. (2015). Performansi Piecewise Polynomial Smooth Support Vector Machine untuk Klasifikasi Desa Tertinggal di Provinsi Kalimantan Timur Tahun 2011. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 7(1), 21–44.

UNVEILING SPATIAL DISPARITIES: EXPLORING HIGH-RISK DIARRHEA AMONG CHILDREN UNDER FIVE USING GEOGRAPHICALLY WEIGHTED QUANTILE REGRESSION

Wara Alfa Syukrilla¹, Yudhie Andriyana², Anneleen Verhasselt³

¹Universitas Islam Negeri Syarif Hidayatullah Jakarta, Tangerang Selatan 15412, Indonesia

²Universitas Padjadjaran, Sumedang 45361, Indonesia

³Universiteit Hasselt, Hasselt 3590, Belgium
e-mail: wara.alfa@uinjkt.ac.id

Received: April 26, 2023; Revised: November 16, 2023; Accepted: December 27, 2024; Published: December 31, 2023

Abstrak

Penelitian ini menyelidiki pengaruh persentase akses air bersih, persentase kebiasaan mencuci tangan, dan kategorisasi toilet sehat pada kuantil atas risiko diare balita di Kota Bandung, Indonesia, menggunakan model Geographically Weighted Quantile Regression pada persentil ke-75 ($\tau = 0,75$). Bandwidth optimal dipilih menggunakan validasi silang. Hasil penelitian menunjukkan bahwa signifikansi, kekuatan, dan arah hubungan antara diare dan faktor risikonya tergantung pada lokasinya. Pada kuantil atas $\tau = 0,75$ Kecamatan Panyileukan diprediksi memiliki risiko diare tertinggi. Di kabupaten ini, ketiga prediktor berpengaruh signifikan terhadap risiko diare pada balita, dengan variabel persentase rumah yang mempraktikkan kebiasaan cuci tangan adalah variabel paling besar pengaruhnya dalam menurunkan risiko diare. Kesimpulannya, akses air bersih, kebiasaan cuci tangan, dan kategori toilet merupakan faktor risiko potensial terjadinya diare pada anak risiko tinggi. Metode GWQR memungkinkan pembuat keputusan untuk menangani masalah diare dengan tepat berdasarkan prediktor mana yang memiliki pengaruh besar pada daerah tertentu yang diminati. Selain itu, GWQR dapat digunakan untuk menyelidiki efek dari berbagai strategi intervensi dan secara efektif mengalokasikan sumber daya terbatas yang tersedia sesuai lokasi yang paling membutuhkannya.

Kata Kunci: Geographically Weighted Quantile Regression, balita, diare

Abstract

We investigate the impact of the percentage of clean water access, the percentage of handwashing habits, and the toilet category factors on the upper quantile of toddlers' diarrhea risks in Bandung City, Indonesia, using the Geographically Weighted Quantile Regression model on the 75th percentile ($\tau=0.75$). The optimum bandwidth was selected using cross-validation. The results show that the significance, strength, and direction of the relationship between diarrhea and its risk factors depend on the location. At the upper quantile $\tau = 0.75$, the Panyileukan district is predicted to have the highest diarrhea risk. In this district, all three predictors significantly affect the toddlers' diarrhea risk, with the variable of the percentage of houses practicing hand washing habits observed to reduce diarrhea risk the most. In conclusion, clean water access, handwashing habits, and toilet category are the potential risk factors for high-risk childhood diarrhea. This method is powerful as it would allow the decision-maker to handle the diarrhea problem aptly by focusing on the predictor that has a significant impact on a particular district of interest. And it can be used to investigate the effect of various intervention strategies and effectively allocate the limited available resources according to the most important locations.

Keywords: Geographically Weighted Quantile Regression, toddlers, diarrhea

INTRODUCTION

Most people think that diarrhea is not a serious situation since almost everyone experienced diarrhea at least once in a lifetime and usually, one recovers within one or two days. Although diarrhea is a treatable and preventable disease, it is the leading killer of toddlers worldwide (UNICEF, 2018). UNICEF has recorded that every day, around 2,000 children in the world die before the age of 5 due to diarrhea (Centers for Diseases Control and Prevention (CDC), 2018). Children residing in low- and middle-income countries, such as Indonesia - ranked among the top 15 countries with the highest number of child deaths caused by pneumonia and diarrhea worldwide - are particularly vulnerable to contracting diarrhea (International Vaccine Access Center (IVAC) and Johns Hopkins Bloomberg School of Public Health, 2018, 2020). Besides threatening the life of toddlers, repeated and frequent diarrhea occurrence could degrade children's growth and cognitive development, as well as increase their vulnerability to other infectious diseases (Liu et al., 2012).

Several past studies show that there is spatial dependency in childhood diarrhea occurrence (Kandala et al., 2007; Azage et al., 2015; Bogale et al., 2017). Most diarrhea is caused by infectious bacteria which easily found in contaminated water or soil. When individuals in a particular area become infected with the disease, they have the potential to pass it on to people in neighboring regions through direct or indirect contact, polluted water sources, or communal facilities. The ease of bacteria transmission through water and soil makes diarrhea could spread from one district to the neighboring districts.

In addition, diarrheal diseases are frequently linked to inadequate sanitation, polluted water sources, and suboptimal hygiene practices (Yilgwan and Okolo, 2012). Shared environmental conditions among neighboring areas can amplify the likelihood of diarrhea outbreaks (Alexander et al., 2018). For instance, the absence of clean water and proper sanitation facilities in one area not only elevates the risk of diarrhea within that area but also poses a heightened threat to nearby regions. Due to these spatial dependencies and

influences, it is crucial to incorporate spatial effects into disease modeling studies.

If spatial effects is not accounted for, it can result in biased and inefficient estimates (Stakhovych et al., 2012). Typically, the spatial approach used assumes spatial stationarity, meaning that the impact of predictors on the response is consistent throughout the study area (Fotheringham et al., 2003). However, this approach may not always be adequate to represent the true variation in real-life situations. This is because the influence of a particular predictor can be location-dependent, meaning that it may be significant in one area but not in another. In cases where the data cannot be fully explained by a single "global" spatial model, it is necessary to consider a Geographically Weighted Regression (GWR) model, which incorporates varying coefficients based on location (Brunsdon et al., 1996). The use of locally spatial analysis, such as GWR, gives the opportunity to detect where the key areas across the large study areas are and detailed information on how interventions affect the hotspot areas of the disease.

Past studies have utilized GWR for considering the spatially varying relationship when analyzing diarrhea and its risk factors (Khoirunnisa et al., 2019; Dunn et al., 2020; Carrel et al., 2011). However, GWR concentrates on modeling the mean of the response. While it can be happened that the upper or lower part of the response responds the predictors differently from the middle part. Our study is more interested in analyzing the upper quantile rather than middle or lower quantile of the response distribution because assessing the risk factors of high-risk diarrhea is more essential in the effort of reducing diarrhea issue. Therefore, this research utilizes a method called Geographically Weighted Quantile Regression (GWQR), which combines Geographically Weighted Regression and a quantile regression scheme developed by Koenker and Bassett in 1978 (Koenker and Bassett, 1978).

Several studies took advantage of quantile regression for exploring the association of several risk factors by subgroups of their response distribution (Beyerlein et al., 2010; Hu et al., 2021). Furthermore, quantile

regression is robust to outliers which overcome the drawback of ordinary GWR that is not robust to outliers (Andriyana et al., 2014). In If the response distribution is skewed, a regression based on the mean may give an incorrect estimate of the impact of predictors on the tail probability of the response. However, quantile regression is less affected by the tail behavior of the response distribution, so it is less likely to over or underestimate the effect of predictors (Alsayed et al., 2020).

The GWQR method takes into consideration both the spatial variability and the distributional differences of the response. In this study, we aim to investigate the impact of the selected factors on the 0.75 quantile of toddlers' diarrhea risks in Bandung, Indonesia using Geographically Weighted Quantile Regression model.

METHODOLOGY

1. Reference Review

In this section, we present a brief description about GWQR. For more details about GWQR regression coefficients, its standard error, and the evaluation of spatial non-stationarity, the reader is referred to (Chen et al., 2012). In the classical Geographically Weighted Regression (GWR), we consider the model

$$Y = \mathbf{X}^T \boldsymbol{\beta}(u, v) + \varepsilon, \quad (1)$$

where Y is the response, $\mathbf{X} = (1, X_1, \dots, X_p)$ is the vector of the predictors with a constant 1 for intercept. $\boldsymbol{\beta}(u, v) = (\beta_0(u, v) + \dots + \beta_p(u, v))^T$ denotes the regression coefficients that are computed for a particular location identified by its geographical coordinates (u, v) . In this study, the geographical coordinates (u, v) identify the locations of various subdistricts within Bandung city. For each set of coordinates (u, v) used in this study, u and v correspond to the latitude and the longitude of the center point of each subdistrict in Bandung city. Additionally, there is a random error term ε , which follows a normal distribution with a mean of zero and a common variance of σ^2 .

Hence $E(Y|\mathbf{X}, u, v) = \mathbf{X}^T \boldsymbol{\beta}(u, v)$. While in the quantile regression (QR) scheme, it is the τ th quantile of ε given X, u, v is equal to zero ($Q_\varepsilon(\tau|\mathbf{X}, u, v) = 0$) since there is no distributional assumption regarding the error term ε . Hence, $Q_Y(\tau|\mathbf{X}, u, v) = \mathbf{X}^T \boldsymbol{\beta}^\tau(u, v)$ (Koenker, 2005).

This study uses GWQR, which is the combination of GWR and QR. The GWQR's parameters are estimated through the process of minimizing the weighted quantile loss function given below (Chen et al., 2012)

$$\sum_{i=1}^n \rho_\tau \left[Y_i - \mathbf{X}_i^T \boldsymbol{\beta}(u_0, v_0) \right] K \left(\frac{d_{i0}}{h} \right), \quad (2)$$

where $\rho_\tau(e)$ is a V-shaped piecewise linear check-function that assign weight τ to the positive residuals and assign weight $(\tau - 1)$ to negative residuals. $K \left(\frac{d_{i0}}{h} \right)$ is the spatial weight. $\mathbf{X}_i^T = (1, X_{1i}, \dots, X_{pi})$ and $i = 1, \dots, n$ is the index for n independent observations. The vector of parameter $\boldsymbol{\beta}(u_0, v_0)$ is the regression coefficients at the location coordinates (u_0, v_0) . The equation Eq. 2 does not have explicit form hence it is solved using linear programming (Chen and Wei, 2005; Koenker, 2005).

2. Analysis Method

We conducted the spatial analysis using GWQR model at the district level at the upper quantile since we are interested in the association at high diarrhea risk. There is no specific theoretical limit for the level of diarrhea risk at which it can be considered high, hence in this paper we use conventional cut off i.e. high risk of diarrhea is 75th percentile ($\tau = 0.75$).

Initial step before applying GWQR model is to check the presence of outliers and spatial heterogeneity, where the spatial heterogeneity detection can be done using the Breusch-Pagan Test ('Afifah et al., 2017; Fitriani and Jaya, 2020; Octaviany et al., 2017; Rahayu et al., 2023; Hothorn et al., 2017). A Breusch-Pagan Test result with a p-value below 0.05 signifies the presence of spatial heterogeneity in the study area, meaning that a predictor's effect varies across locations and the use of the global spatial



Figure 1. Maps of 30 districts in Bandung city, Indonesia

regression model is insufficient, highlighting the need for a geographically weighted regression model.

The next step is to generate spatial weight. To allocate a spatial weight to each observation, a kernel method is used, which employs a kernel function $K(\cdot)$ that considers two factors: the distance (d_{i0}) between the coordinates of the i th observation (u_i, v_i) and the regression point (u_0, v_0), measured using Euclidean distance, and a bandwidth parameter (h). The closer an observation is to the regression point (u_0, v_0), the greater weight it will be assigned, while those further away will have a lower weight. The bandwidth parameter ($h \geq 0$) determines the degree of distance-decay in the weighting process and to adjust the smoothness of the resulting coefficients.

This study uses a bi-square kernel with adaptive bandwidth where the kernel size will be big/small following the density of observations at a region. The bi-square kernel was preferred due to its robustness in handling data with outliers and non-normal characteristics, such as skewness. Given that our diarrhea occurrence dataset exhibits skewness, the bi-square kernel was deemed suitable for mitigating the influence of outliers and providing reliable density estimates in the

presence of non-well-behaved data distributions. Fotheringham et al. (2003) suggested that different kernels produce similar results, but varying bandwidths can lead to different outcomes. Therefore, it is crucial to choose the appropriate bandwidth. The paper uses the Leave One Out Cross Validation (CV) criteria to determine the best bandwidth, where the bandwidth with the lowest CV score will be selected. All of the analyses were done in R software version 4.0.0. The GWQR analysis in this thesis utilized a modified version of GWmodel package sourced from Khaeri (2018).

Data and Variables

According to the Indonesian Ministry of Health (2017) the West Java province has the highest incidence of diarrhea in Indonesia, and among the top five cities in the province with the highest incidence of diarrhea is Bandung, the provincial capital (West Java Department of Health, 2017). As a result, serious measures need to be taken in Bandung to control and treat diarrhea in order to reduce its occurrence. This study focuses on the diarrhea cases in children under five in Bandung city, Indonesia. Bandung has 30 districts with the geographical maps presented in Figure 1.

Raw data for the response variable were generated from the Bandung Department of Health, Indonesia and contains information of the number of toddler's diarrhea occurrences in 2015 at $n = 30$ districts of Bandung city together with the number of toddlers in each district. The response of interest, namely diarrhea risk, was calculated by dividing the number of cases of diarrhea in each district by the number of toddlers in that district, and then multiplying the result by 100. This calculation produces a range of values between 0 and 100. The derived response data are available on request to the corresponding author.

The three predictors used in this study are thought to be linked with the occurrence of diarrhea infection, according to the diarrhea bulletin released by the Indonesian Ministry of Health (2011) and the United Nations' WASH (water, sanitation, and hygiene) campaign (UNICEF, 2019). These three predictors have a range of 0-100 and are obtained from the Bandung Department of Health. The first predictor (X_1) is the percentage of households in each district that utilize clean water for their everyday tasks, the second predictor (X_2) is the percentage of households in every district that

practicing hand washing habit in their daily activities, and the third predictor is the percentage of houses in every district that have toilets fulfilling the healthy toilet criteria by Indonesian Ministry of Health (Indonesian Ministry of Health, 2014).

RESULTS

Exploratory Data Analysis

We present a boxmap in Figure 2 which is a box plot of map data. Boxmap helps to detect the presence of outliers in a response variable through a geographical map visualization. Figure 2 shows that there are no outliers present in the response.

The histogram of the response variable in Figure 3 reveals that the distribution of diarrhea risk is not symmetrical. The skewed pattern of the response is suitable to be handled with a quantile regression scheme since quantile regression does not make any assumptions about the distribution of the response variable. While if the skewed pattern of the response is handled using mean-based regression, it could cause violation of regression assumption.

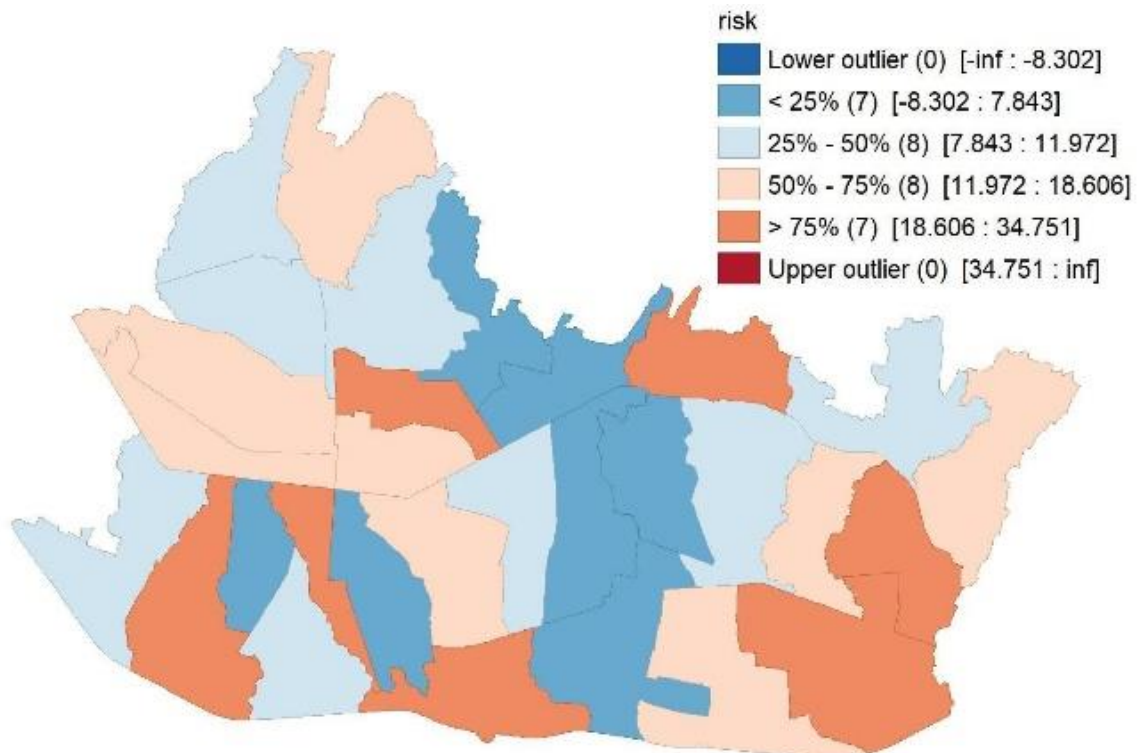


Figure 2. Boxmap of the observed diarrhea risk

After considering the spatial effect, the Breusch-Pagan test was conducted to examine spatial heterogeneity. The test yielded a p-value of 0.043, which is significant at a level of $\alpha = 0.05$, indicating that there is spatial heterogeneity in the data. This suggests that the association between the predictors and the response variable varies across space and that utilizing locally varying spatial models such as the Geographically Weighted model would be appropriate

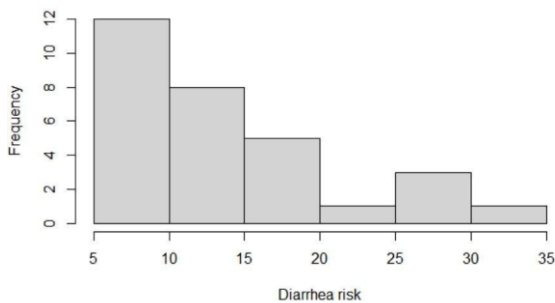


Figure 3. Histogram of the observed diarrhea risk

Displayed in Figure 4 are the scatter plots of the response at the Y-axis and the predictor at the X-axis. Each point represents the risk value of each district. It can be observed that the amount of variability tends to increase with higher values of the predictor variable, as evidenced by the greater distribution of data points at higher predictor values in all three scatter plots.

Regarding the left graph representing X_1 (the percentage of houses with clean water access), the majority of the data points are concentrated at a range of 99 to 100 percent. However, the prevalence of diarrhea ranges from 5 to 32 percent despite this high

proportion of houses with clean water access. An alike case happens with the plot of X_2 at the middle (the percentage of houses practicing hand washing habits). Although the risk of diarrhea spans a wide range from low to high, most districts have a percentage of houses with hand washing habits exceeding 90%. The scatterplot for the third predictor on the right displays a greater range of variability than the previous two predictors, as indicated by the spread of points across a range of low to high values for X_3 (the percentage of houses with healthy toilets status). Additionally, it was noted that one district has low values for clean water percentage, hand washing habits, and healthy toilets, yet the risk of diarrhea in that district is also low.

Results of Geographically Weighted Quantile Regression

The adaptive bandwidth method was used with the GWQR bi-square kernel, and it was discovered through cross-validation that the best bandwidth for the 0.75th quantile is $h = 28$.

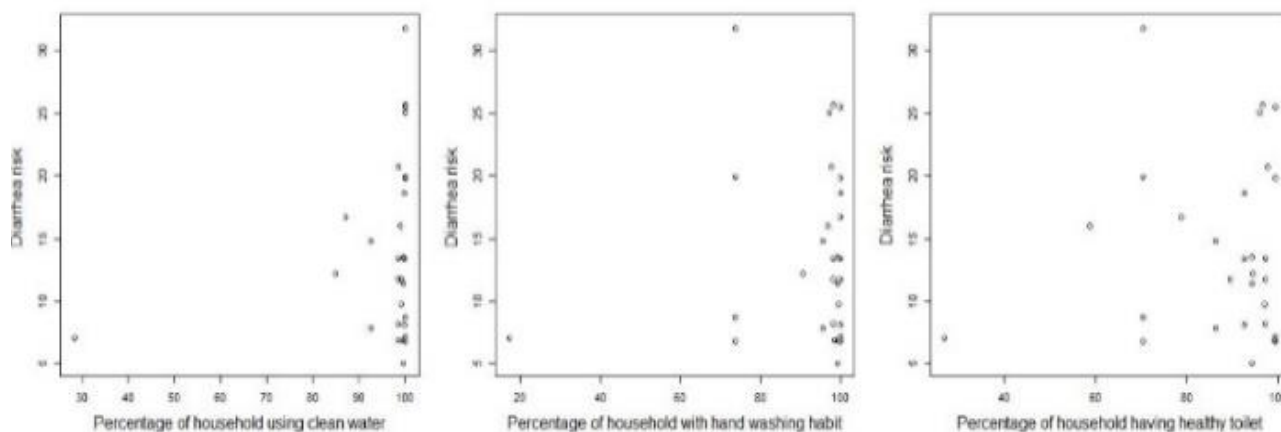


Figure 4. Scatter plot of predictors (left to the right: X_1 - X_3) against the response

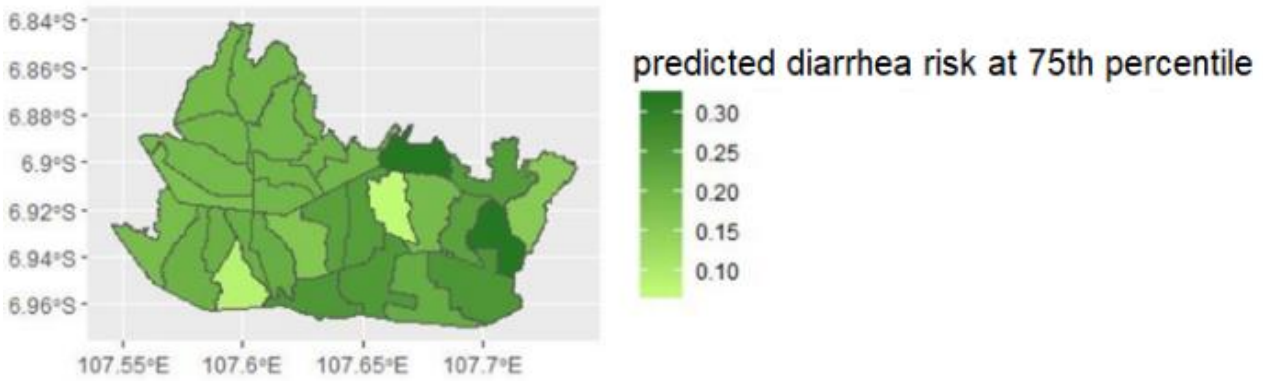


Figure 5. GWQR predictions for diarrhea risk at the 75th percentile

Figure 5 illustrates the map of diarrhea risk predictions from GWQR model based on the 0.75 quantile. The darker colors in the eastern part of the map signify two specific districts that have been identified as having the highest probability of experiencing diarrhea. This is in comparison to all other districts in the upper percentile, with a diarrhea risk of over 30%. In these two districts, 25 out of every 100 children under the age of five are estimated to have a 30% or greater risk of contracting diarrhea. As a result, it is recommended that the health authorities give particular attention to the issue of diarrhea in these districts.

Figure 6, 7, and 8 display the maps for the GWQR parameter estimates and their corresponding significance tests at the upper quantiles ($\tau = 0.75$). The degree of darkness in the color on the maps represents the magnitude of the positive impact of a covariate on the response variable. Conversely, lighter shades suggest that the effect of a predictor is either slightly positive or negative. The areas

on the maps that are colored white are locations where the predictor's effect is not significant at a 5% level of significance.

The association between the percentage of houses that use clean water (X_1) and diarrhea risk is depicted in Figure 6. The western section of the map shows a negative effect of the percentage of clean water on diarrhea risk at the 75th percentile, indicating that the higher percentage of houses that use clean water for daily activities, the lower toddlers' diarrhea risk. However, this effect is not significant in one district. Furthermore, while the impact of the percentage of houses with clean water (X_1) on the incidence of diarrhea is statistically significant in nearly all districts, the extent of this impact differs across various locations. Its effect gets larger as we move from the west to the east areas of Bandung. These results illustrate how spatial non-stationarity exists in the data, where the effect of a predictor depends on where it is evaluated.

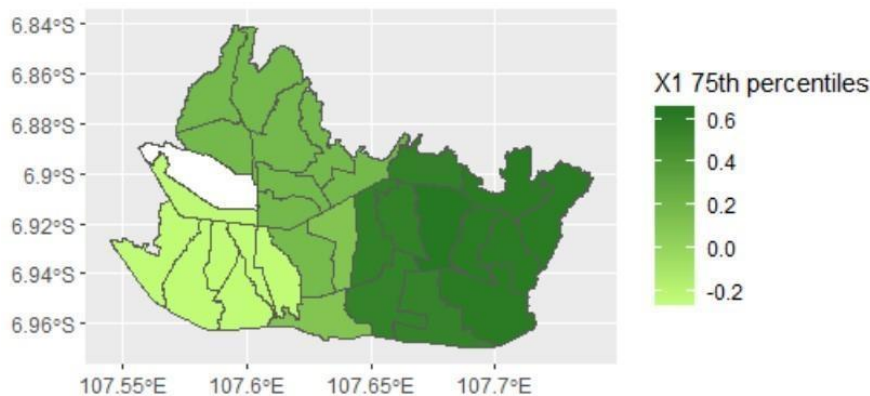


Figure 6. Maps of GWQR $\tau = 0.75$ Estimates for X_1 (Percentage of Houses with Clean Water Usage) and its significance at $\alpha = 0.05$

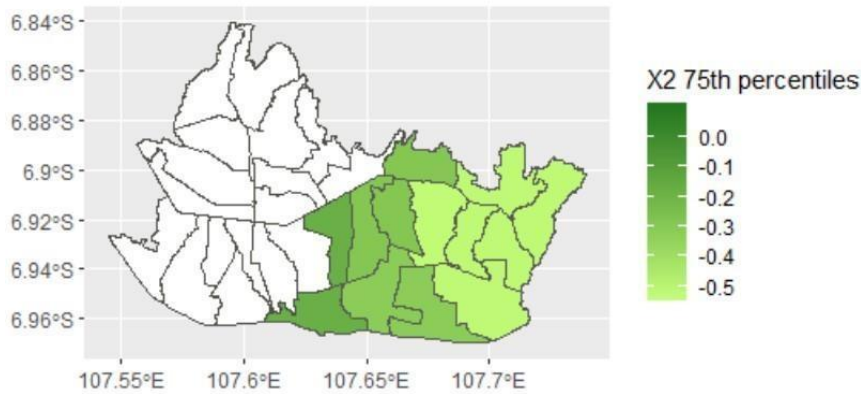


Figure 7. Maps of GWQR $\tau = 0.75$ Estimates for X_2 (Percentage of houses practicing hand washing habit) evaluated at the 5% significant level

The map shown in Figure 7 displays the significance of parameter estimates for the percentage of houses that practiced hand washing habits (X_2) for all districts at the 0.75th quantile level. At the 75th percentile, the percentage of houses that practiced hand washing habits is a significant factor in reducing the likelihood of toddlers in the eastern half of Bandung developing diarrhea. However, this predictor was observed to have a non-significant effect at the other half of Bandung area. These results are valuable as it helps the decision makers to efficiently identify which districts will experience a decreasing diarrhea risk when the government promotes policy or campaign related to handwashing habits.

Figure 8 illustrates the geographic distribution of GWQR coefficients for X_3 , which represents the percentage of houses with healthy toilets status significant at a 5% significance level. The study found that the percentage of houses with healthy toilets (X_3) is a significant predictor of the risk of toddlers

developing diarrhea in both the western and eastern regions of Bandung when at the 75th percentile. However, it is not a significant predictor in the central part of the city. Again, this illustrates that the effect of predictors can be different between locations, both in the significance and magnitude.

All three predictors were found to have significant influences on toddlers' diarrhea in the following districts: Gede Bage, Ujung Berung, Cinambo, Batu Nunggal, Bandung Kidul, Cibiru, Panyileukan, and Arcamanik. These districts are mostly located in the eastern area of Bandung city maps picture.

There was no location where no significant predictor variables influenced the diarrhea events. At least one predictor significantly influences diarrhea risks in Bandung districts. For example, diarrhea risks in the Cicendo district were influenced by X_3 (Percentage of Houses with Healthy Toilet). Diarrhea events at the following locations were influenced by only X_1 (Percentage of Houses with Clean Water Usage): Sukajadi, Cidadap,

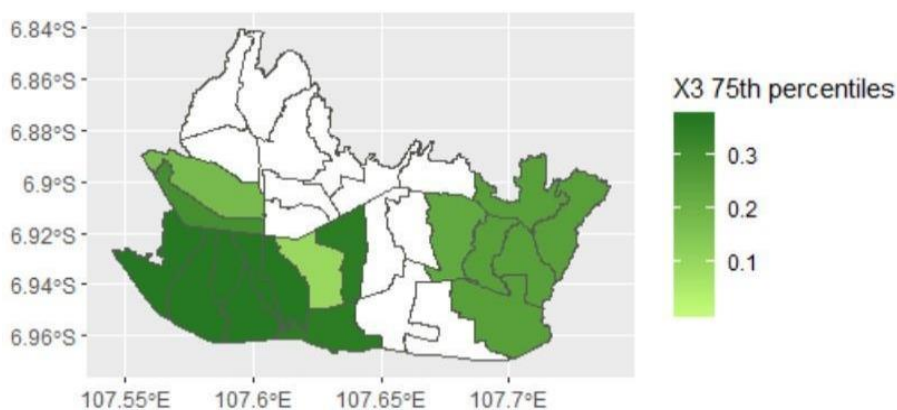


Figure 8. Maps of GWQR $\tau = 0.75$ Estimates for X_3 (Percentage of Houses with Healthy Toilet) significant at $\alpha = 5\%$

Coblong, Bandung Wetan, Sumur Bandung, Cibeunying Kidul, Cibeunying Kaler, Sukasari.

DISCUSSION

The use of quantile regression scheme is beneficial over the mean-based model in that it can handle the skewed data in this study which might have violated assumptions in mean-based regression models. It also allows us to characterize the impact of predictors on subgroups of response distribution using all of the available data. In this research, the Geographically Weighted Quantile Regression (GWQR) model was utilized to address the research questions regarding the association between selected predictors and the risk of diarrhea in children under the age of 5 in various spatial locations. The evaluation was done at the upper quantile ($\tau = 0.75$) of the diarrhea risk distribution. This method is powerful as it can reflect spatial variation of targeted childhood high risk diarrhea and identify the important risk factors at every district especially in Bandung where the water access and sanitation are varying between districts.

A lot of households in Bandung obtain their water supply either from the Municipal Water Corporation or from wells. The Municipal Water Corporation divides its water distribution service and pipeline systems into three sub-regions for those subscribed to it. Unequal access to water across different districts of Bandung poses a varying risk of diarrhea. Additionally, there are still some households without septic tanks for their toilets in Bandung, which prevents the city from achieving complete ODF status (Bandung Public Relations, 2020). They flowed the feces disposal to the river or practiced open defecation. The evidence could imply that regions situated close to rivers or where waste is disposed are more susceptible to soil and groundwater pollution, resulting in an unstable diarrhea risk in various districts.

Our results show that the location that is predicted to have the highest diarrhea risk in Bandung is the district at the coordinate (6.93° S, 107.7° E) named Panyileukan district. Using GWQR we can identify toddlers' diarrhea

determinants locally for every district differently. As an example, let us compare two districts namely Panyileukan and Antapani districts, where the former is the district that predicts the highest diarrhea risk while the later predicts the smallest diarrhea risk at the 75th percentile of the response distribution. In the Panyileukan district, the covariate X_3 or the percentage of houses with healthy toilets is considered as an important predictor as its effect is significant in this district. On the other hand, in Antapani district, the predictor X_3 was observed to have no effect. Looking at the X_2 predictor, although its effect is significant at both locations, X_2 is predicted to have a bigger reduction on diarrhea risks compared to it in the Antapani district.

The findings indicate that certain coefficients exhibit a negative association between the predictors and the response, while others indicate a positive association. The positive coefficients may appear counterintuitive because it is not anticipated that an increase in the proportion of households with access to clean water, hygienic toilets, and handwashing practices would amplify the risk of diarrhea in young children. After examining the scatter plot in Figure 4, it can be observed that the diarrhea risk values (Y) vary across districts, ranging from low to high values. However, the predictors, such as the percentage of households with clean water access, handwashing habits, and healthy toilets, have high percentages across most districts, irrespective of the level of diarrhea risk. The percentages of X_1 and X_2 , in particular, exceed 98% in almost all districts, even those with a very high risk of diarrhea. Furthermore, a correlation seems to exist between X_3 and Y , with the response increasing as X_3 increases. Additionally, there is one district with low percentages in all predictors but still has a low diarrhea risk. These patterns cause the model to predict a positive linear relationship between the predictors and the response.

In addition, since the number of data used in this study is 30 and the weighting function is adaptive, thus less data was used to model a regression point than 30. This limitation of the study might contribute to the

counterintuitive positive associations between predictors and response in some locations. Therefore, future studies are suggested to use a larger number of data points or opt for another kernel choice when modeling using GWQR.

CONCLUSIONS

In conclusion, clean water access, handwashing habits, and toilet category are the potential risk factors of high risk childhood diarrhea. The significance, strength, and direction of the effect varies between districts. The estimation of GWQR at the 75th percentile predicts the district with highest risk is the one at the coordinate (6.93° S, 107.7° E) namely Panyileukan district. In this district, all of the three predictors significantly affect the toddlers' diarrhea risk but the variable percentage of houses practicing hand washing habit (X_2) is observed to reduce diarrhea risk the most. This information is important and would enable the decision-maker to effectively address the issue of diarrhea by identifying which predictor has a significant impact on a particular district of concern. At an expanded level, GWQR can be used to investigate the effect of various intervention strategies, and effectively allocate the limited available resources according to which locations are most important. This can also be applied to other cities in Indonesia and other high burden countries to reduce the world diarrhea number, and a large number of toddlers' lives would be saved.

REFERENCES

- ‘Afifah, Rawyanil, Yudhie Andriyana, and IGN Mindra Jaya. 2017. Robust geographically weighted regression with least absolute deviation method in case of poverty in Java Island . AIP Conference Proceedings 1827(1), 020023.
- Alexander, Kathleen A, Alexandra K Heaney, and Jeffrey Shaman. 2018. Hydrometeorology and flood pulse dynamics drive diarrheal disease outbreaks and increase vulnerability to climate change in surface-water-dependent populations: A retrospective analysis . PLOS Medicine 15(11), e1002688.
- Alsayed, Ahmed RM, Zaidi Isa, Sek Siok Kun, and Giancarlo Manzi. 2020. Quantile Regression to Tackle the Heterogeneity on the Relationship Between Economic Growth, Energy Consumption, and CO2 Emissions . Environmental Modeling and Assessment 25(2), 251–58.
- Andriyana, Y, I Gijbels, and A Verhasselt. 2014. P-splines quantile regression estimation in varying coefficient models . Test 23(1), 153–94.
- Azage, Muluken, Abera Kumie, Alemayehu Worku, and Amvrossios C Bagtzoglou. 2015. Childhood Diarrhea Exhibits Spatiotemporal Variation in Northwest Ethiopia: A SaTScan Spatial Statistical Analysis . PLoS ONE 10(12), 1–18.
- Bandung Public Relations. 2020. ODF di Kota Bandung Terus Membaik membaik [ODF (Open Defecation Free) in Bandung Continues to Improve]. <https://humas.bandung.go.id/layanan/odf-di-kota-bandung-terus-membaik>.
- Beyerlein, A, AM Toschke, and R Von Kries. 2010. Risk factors for childhood overweight: Shift of the mean body mass index and shift of the upper percentiles: Results from a cross-sectional study . International Journal of Obesity 34(4), 642–48.
- Bogale, Getahun Gebre, Kassahun Alemu Gelaye, Degefie Tibebe Degefie, and Yalemzewod Assefa Gelaw. 2017. Spatial patterns of childhood diarrhea in Ethiopia: Data from Ethiopian demographic and health surveys (2000, 2005, and 2011) . BMC Infectious Diseases 17(1), 1–10.
- Brunsdon, Chris, A Stewart Fotheringham, and Martin E Charlton. 1996. Geographically Weighted Regression . Geographical Analysis 28(4), 281–98.
- Carrel, Margaret, Veronica Escamilla, Jane Messina, Sophia Giebultowicz, Jennifer Winston, Mohammad Yunus, P Kim Streatfield, and Michael Emch. 2011. Diarrheal disease risk in rural Bangladesh decreases as tubewell density increases: A zero-inflated and geographically weighted analysis .

- International Journal of Health Geographics 10, 1–9.
- Centers for Diseases Control and Prevention (CDC). 2018. *Diarrhea: Common Illness, Global Killer*.
- Chen, Colin, and Ying Wei. 2005. Computational issues for quantile regression . *Sankhya: The Indian Journal of Statistics* 67(2), 399–417.
- Chen, Vivian Yi Ju, Wen Shuenn Deng, Tse Chuan Yang, and Stephen A Matthews. 2012. Geographically Weighted Quantile Regression (GWQR): An Application to U.S. Mortality Data . *Geographical Analysis* 44(2), 134–50.
- Dunn, Gillian, Glen D Johnson, Deborah L Balk, and Grace Sembajwe. 2020. Spatially varying relationships between risk factors and child diarrhea in West Africa, 2008-2013 . *Mathematical Population Studies* 27(1), 8–33.
- Fitriani, Rinda, and I Gede Nyoman Mindra Jaya. 2020. Spatial modeling of confirmed COVID-19 pandemic in East Java province by geographically weighted negative binomial regression . *Commun. Math. Biol. Neurosci.* 2020(0), Article ID 58.
- Fotheringham, A Stewart, Chris Brunsdon, and Martin Charlton. 2003. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: John Wiley & Sons.
- Hothorn, Torsten, Achim Zeileis, and Roger Bivand. 2017. *bptest.sarlm: Breusch-Pagan test for spatial models*. <https://rdrr.io/rforge/spdep/src/R/bptest.sarlm.R>.
- Hu, Liangyuan, Jiayi Ji, Yan Li, Bian Liu, and Yiyi Zhang. 2021. Quantile Regression Forests to Identify Determinants of Neighborhood Stroke Prevalence in 500 Cities in the USA: Implications for Neighborhoods with High Prevalence . *Journal of Urban Health* 98(2), 259–70.
- Indonesian Ministry of Health. 2011. Situasi Diare di Indonesia (Diarrhea Situation in Indonesia) . *Jurnal Buletin Jendela Data & Informasi Kesehatan* 2, 1–44.
- KEMENKES. 2014. Peraturan Menteri Kesehatan Republik Indonesia Nomor 3 Tahun 2014 tentang Sanitasi Total Berbasis Masyarakat. Indonesia.
- KEMENKES. 2017. *Profil Kementerian Kesehatan Indonesia 2017 (Indonesian Health Profile in 2017)*. Jakarta.
- International Vaccine Access Center (IVAC), and Johns Hopkins Bloomberg School of Public Health. 2018. *Pneumonia & diarrhea progress report 2018*.
- International Vaccine Access Center (IVAC), and Johns Hopkins Bloomberg School of Public Health. 2020. *Pneumonia and Diarrhea Progress Report 2020*.
- Kandala, Ngianga-bakwin, Ji., Nigel Chen Stallard, Saverio Stranges, and Francesco P Cappuccio. 2007. Spatial Analysis of Risk Factors for Childhood Morbidity in Nigeria . *The American Society of Tropical Medicine and Hygiene Spatial* 77(4), 770–78.
- Khaeri, Hirlan. 2018. *GEOGRAPHICALLY WEIGHTED QUANTILE REGRESSION PADA ANALISIS ANGKA HARAPAN HIDUP PENDUDUK INDONESIA* [Unpublished doctoral or master’s thesis or dissertation]. Universitas Padjajaran.
- Khoirunnisa, Anis, I Gede Nyoman, and Mindra Jaya. 2019. Modelling and Mapping The Average of Incidence Rate Diarrhea Among Toddlers in Bandung City in 2013-2018 Using Geographically Weighted Regression . *International Journal of Innovative Science, Engineering & Technology* 6(11), 115–19.
- Koenker, Roger. 2005. *Quantile Regression (Econometric Society Monographs; No. 38)* . Nature Methods. Vol. 16. New York: Cambridge University Press.
- Koenker, Roger, and Gilbert Bassett. 1978. Regression Quantiles . *Econometrica* 46(1), 33.
- Liu, Li, Hope L Johnson, Simon Cousens, Jamie Perin, Susana Scott, Joy E Lawn, Igor Rudan, Harry Campbell, Richard Cibulskis, Mengying Li, Colin Mathers, and Robert E Black. 2012. Global, regional, and national causes of child mortality: an updated systematic

- analysis for 2010 with time trends since 2000 . The Lancet 379(9832), 2151–61.
- Octaviany, Toni Toharudin, and IGN Mindra Jaya. 2017. Geographically weighted poisson regression semiparametric on modeling of the number of tuberculosis cases (Case study: Bandung city) . AIP Conference Proceedings 1827(1), 020022.
- Rahayu, Azka Larissa, Gumgum Darmawan, and I Gede Nyoman Mindra Jaya. 2023. Calculation of the Risk Index for Diarrhea, ISPA, and Pneumonia in Toddlers in the City of Bandung Using Geographically Weighted Principal Component Analysis . Indonesian Journal of Advanced Research 2(4), 285–300.
- Stakhovych, Stanislav, Tammo HA Bijmolt, and Michel Wedel. 2012. Spatial Dependence and Heterogeneity in Bayesian Factor Analysis: A Cross-National Investigation of Schwartz Values . Multivariate Behavioral Research 47(6), 803–39.
- UNICEF. 2018. *Every child survives and thrives Global Annual*.
- UNICEF. 2019. *Water under Fire: For every child, water and sanitation in complex emergencies*.
- West Java Department of Health. 2017. *Profil Kesehatan Dinas Kesehatan Provinsi Jawa Barat 2017 (The Health Profile of West Java Province in 2017)* . Diskes Jabarprov.
- Yilgwan, CS, and SN Okolo. 2012. Prevalence of diarrhea disease and risk factors in Jos University Teaching Hospital, Nigeria . Annals of African Medicine 11(4), 217–21.

PENINGKATAN KUALITAS STATISTIK RESMI PRODUKTIVITAS PADI MELALUI IMPUTASI DATA NON-RESPONS MENGGUNAKAN MODEL ADITIF GEOSPASIAL

Muhlis Ardiansyah^{1,2}

¹BPS Kabupaten Kotawaringin Timur, Kalimantan Tengah, Indonesia

² Departemen Statistika, IPB University, Kampus IPB Dramaga, Bogor, Indonesia
e-mail: muhli@bps.go.id; muhlisardiansyah@apps.ipb.ac.id

Received: November 15, 2022; Revised: October 11, 2023; Accepted: November 21, 2023; Published: December 31, 2023

Abstrak

Penelitian ini dilatarbelakangi oleh masalah non-respons pada Survei Ubinan yang diselenggarakan oleh Badan Pusat Statistik (BPS) sebagai penyedia statistik resmi. BPS memiliki visi sebagai penyedia data statistik berkualitas untuk Indonesia maju. Penanganan non-respons sangat penting untuk mendukung visi tersebut karena non-respons berpotensi menyebabkan beberapa karakteristik sampel menjadi tidak terwakili. Penelitian ini mengusulkan teknik imputasi data non-respon melalui pemodelan statistik. Model yang diusulkan adalah model aditif dengan penambahan fungsi pemulusan geospasial *thin plate regression splines* (TP) dan *Gaussian process* (GP). Pemilihan model terbaik berdasarkan *Mean Squared Error of Prediction* (MSEP) terkecil dari 1000 iterasi. Kemudian dibandingkan rata-rata produktivitas padi antara teknik mengabaikan non-respons (*listwise deletion*) dan imputasi melalui tiga skenario data non-respons. Hasil penelitian menunjukkan bahwa model dengan penambahan fungsi pemulus GP memberikan kinerja terbaik dengan MSEP tekecil. Hasil lainnya menunjukkan bahwa metode imputasi data non-respons lebih baik dibandingkan dengan mengabaikan non-respons. BPS dapat mempertimbangkan metode imputasi untuk meningkatkan kualitas data statistik resmi produktivitas padi.

Kata Kunci: imputasi, geospasial, non-respons, statistik resmi

Abstract

This study is motivated by the non-response problem in the Crop Cutting Survey conducted by the BPS-Statistics Indonesia as the official statistics provider. BPS has a vision of providing quality statistical data for advanced Indonesia. Handling non-response is essential to supporting this vision because non-response can potentially cause some sample characteristics to be unrepresented. This study proposed a non-response data imputation technique through statistical modeling. The proposed model was an additive model with the addition of geospatial smoothing functions of thin plate regression splines (TP) and Gaussian process (GP). Selection of the best model based on the smallest MSEP of 1000 iterations. Then we compared the average rice productivity between listwise deletion and imputation techniques through three scenarios of non-response data. The results showed that the model with the addition of the GP smoothing function gave the best performance with the smallest MSEP. The other results showed that the imputation method of non-response data is better than ignoring non-response. BPS can consider the imputation method to improve the quality of official statistics on rice productivity.

Keyword: imputation, geospatial, non-response, official statistics

PENDAHULUAN

Badan Pusat Statistik (BPS) sebagai lembaga penyedia statistik resmi (*official statistics*) melakukan penghitungan produktivitas tanaman padi (*Oryza sativa*) melalui Survei Ubinan. Data resmi produktivitas padi dibutuhkan untuk mendapatkan informasi tentang capaian target kedua pembangunan berkelanjutan (*Sustainable Development Goals - SDGs*). Data tersebut dihitung dengan mengambil sampel padi yang siap panen pada petak sawah maupun bukan sawah kemudian mengambil secara acak satu plot padi berukuran $2.5 \times 2.5 m^2$ untuk dipanen, dibersihkan, dan ditimbang. Salah satu permasalahan yang dihadapi pada pelaksanaan survei ini adalah terdapat non-respons yang cukup banyak (Ardiansyah *et al.*, 2021).

Data non-respons pada Survei Ubinan disebabkan oleh beberapa hal. Pertama, petugas kadang-kadang tidak berhasil mengambil sampel plot tanaman padi karena sudah dipanen oleh petani terpilih. Kedua, kebijakan pembatasan sosial oleh pemerintah. Ketiga, perubahan cara pemanenan di beberapa petani dari cara tebas ke mesin *combine harvester* sehingga menyebabkan petugas tidak dapat mengambil sampel (Ardiansyah *et al.*, 2020).

Data non-respons menyebabkan berkurangnya keterwakilan sampel sehingga sebaran sampel plot ubinan menjadi tidak merata. Jika sampel yang dipilih berada pada kelompok dengan produktivitas padi rendah maka hasil estimasi akan lebih rendah dari nilai sebenarnya (*underestimate*) dan begitupun sebaliknya. Guna mendukung visi BPS sebagai penyedia data statistik berkualitas untuk Indonesia maju (*provider of qualified statistical data for advanced Indonesia*), penanganan non-respons menjadi penting untuk dilakukan guna meningkatkan kualitas data produktivitas padi.

Terdapat tiga kerugian yang disebabkan adanya data non-respons pada suatu survei. Pertama, berkurangnya kekuatan statistik akibat ukuran sampel yang berkurang. Hal ini berpotensi meningkatkan risiko kesalahan jenis II (*Error type II*) yang merupakan kesalahan akibat menerima H_0 padahal H_1

benar. Kedua, data non-respons berpotensi menimbulkan bias pendugaan. Ketiga, dugaan galat baku menjadi *over* atau *underestimate* (Curley *et al.*, 2019).

Penelitian ini merupakan kelanjutan dari penelitian sebelumnya. Pada penelitian sebelumnya, Ardiansyah & Tofri (2019) mengusulkan metode wawancara pascapanen berdasarkan pengakuan petani untuk mengatasi permasalahan non-respons pada Survei Ubinan. Hasilnya, metode wawancara pasca panen belum cukup bukti untuk menggantikan metode pengukuran plot ubinan karena memberikan hasil produktivitas padi yang lebih rendah (*underestimate*). Oleh karena itu, penelitian ini mengusulkan cara lain yaitu teknik imputasi data non-respons melalui pemodelan statistika.

Secara garis besar, terdapat tiga cara untuk menangani data non-respons pada suatu survei. Pertama, petugas melakukan kunjungan ulang (*revisit*) untuk melengkapi data. Cara ini tidak mungkin dilakukan pada kasus Survei Ubinan karena sampel terpilih sudah terlanjur dipanen oleh petani sehingga tidak mungkin dilakukan pengukuran berat gabah. Cara kedua adalah dengan menghapus atau mengabaikannya (*listwise deletion*) yaitu pendugaan parameter hanya dilakukan dengan menggunakan data respon yang berhasil dikumpulkan dan mengabaikan data non-respons. Cara ketiga adalah dengan mengimputasinya (*imputation*) yaitu menduga data non-respons dengan metode tertentu.

Chhabra *et al.*, (2019) membagi cara imputasi data non-respons menjadi dua. Pertama, cara tradisional seperti imputasi rata-rata, imputasi acak sesuai kelasnya, imputasi hot-deck berurutan, dan imputasi hot-deck hirarki. Kedua, metode modern seperti metode imputasi berbasis regresi, imputasi berganda, dan imputasi menggunakan pembelajaran mesin.

Metode imputasi yang dikembangkan pada penelitian ini adalah metode imputasi berbasis regresi dengan memanfaatkan peubah lainnya. Pada Survei Ubinan, kita dihadapkan pada peubah respon yang sulit diukur dan peubah prediktor yang mudah diperoleh. Peubah prediktor tersebut akan dimanfaatkan untuk mengimputasi data non-respons.

Akurasi dan presisi imputasi data non-respons berbasis regresi dapat ditingkatkan dengan mempertimbangkan berbagai model statistika seperti *Generalized Linear Model* (GLM) (Nelder & Wedderburn, 1972), *Generalized Additive Models* (GAM) (Hastie & Tibshirani, 1986), *Generalized Linear Mixed Models* (GLMM) (Breslow & Clayton, 1993), dan *Generalized Additive Mixed Models* (GAMM) (Lin & Zhang, 1999). Seiring perkembangan teknologi, model dengan pendekatan berbasis data (*data-driven*) untuk mengeksplorasi pola nonlinier dapat ditempuh dengan metode gabungan antara parametrik dan nonparametrik. Pada penelitian ini akan ditambahkan fungsi pemulus geospasial pada model aditif yaitu *thin plate regression splines* (TP) dan *Gaussian process smooths* (GP). Model imputasi yang diajukan ada tiga, yaitu model campuran linier tanpa penambahan fungsi pemulus geospasial (GLMM), model campuran linier dengan penambahan fungsi pemulus TP (GLMM+TP), dan model campuran linier dengan penambahan fungsi pemulus GP (GLMM+GP). Model yang diajukan dipilih karena memiliki fleksibilitas sehingga diduga mampu memberikan akurasi imputasi yang lebih baik dibanding model tanpa penambahan fungsi pemulus geospasial.

Penelitian ini bertujuan untuk memilih metode imputasi terbaik antara model GLMM, GLMM+TP, dan GLMM+GP. Tujuan kedua adalah mengevaluasi kinerja metode imputasi dibandingkan metode mengabaikan atau menghapus data non-respons (*listwise deletion*). Imputasi dilakukan dengan mengganti data non-respons dengan nilai dugaan. Setelah semua data hilang diganti, dataset dianalisis menggunakan teknik standar sebagaimana pendugaan pada data yang lengkap.

METODE

Langkah pertama adalah mengidentifikasi apakah ada autokorelasi spasial pada data produktivitas padi menggunakan indeks Moran. Indeks Moran adalah ukuran autokorelasi spasial yang diusulkan pertama kali oleh Patrick Alfred

Pierce Moran pada tahun 1950. Indeks Moran diformulasikan sebagai berikut:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (y_i - \bar{y})^2}, \dots \dots \dots (1)$$

dengan

y_i = nilai amatan lokasi ke- i ($i = 1, 2, \dots, n$),

y_j = nilai amatan lokasi ke- j ($j = 1, 2, \dots, n$),

\bar{y} = rata-rata nilai amatan dari seluruh lokasi,

w_{ij} = pembobot kedekatan antara lokasi ke- i dengan lokasi ke- j .

Pembobot yang dipilih adalah invers jarak. Informasi mengenai lokasi dilihat dari titik longitude dan latitude. Titik longitude dan latitude terlebih dahulu dinormalisasi. Informasi inilah yang digunakan untuk menghitung jarak antar titik sampel. Kekuatan ketergantungan spasial akan menurun sesuai dengan semakin jauhnya jarak. Jarak ditentukan dengan jarak Euclid. Uji signifikansi Indeks Moran perlu dilakukan untuk melihat nyata tidaknya autokorelasi spasial. Hipotesis untuk Indeks Moran adalah: $H_0 : I = 0$ (Tidak terdapat autokorelasi spasial)

$H_1 : I > 0$ (Terdapat autokorelasi spasial.)

Statistik uji: $Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}} \sim N(0, 1)$

dengan nilai harapan Indeks Moran pada asumsi normal dan acak adalah $E(I) = -\frac{1}{n-1}$;

$Var(I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{(n^2 - 1) S_0^2} - \left[\frac{1}{n-1} \right]^2$; $S_0 =$

$\sum_{i=1}^n \sum_{j=1}^n w_{ij}$; $S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2$;

$S_2 = \sum_{i=1}^n (\sum_{j=1}^n w_{ij} + \sum_{j=1}^n w_{ji})^2$. Kriteria

ujinya adalah tolak H_0 jika $|Z(I)| > Z_\alpha$ (Djuraidah 2020).

Model Aditif Terampat (*Generalized Additive Models/ GAM*)

Model aditif terampat merupakan perpaduan antara peubah prediktor parametrik dan nonparametrik atau yang sering disebut model semiparametrik dengan peubah respon termasuk ke dalam keluarga eksponensial. GAM pertama kali diperkenalkan oleh Hastie dan Tibshirani pada tahun 1986 yang menganggap bahwa mean dari peubah respon yang termasuk dalam keluarga eksponensial tergantung pada prediktor berupa fungsi aditif melalui fungsi hubung (*link function*).

Misalkan $\mathbf{X}^{[j]}$ menunjukkan matriks model dan $\mathbf{S}^{[j]}$ menunjukkan matriks penalti untuk f_j . Kemudian digabungkan antara \mathbf{A} dan $\mathbf{X}^{[j]}$ berdasarkan kolom, untuk membuat model matriks $\mathbf{X} = (\mathbf{A}; \mathbf{X}^{[1]}; \mathbf{X}^{[2]}; \dots)$. $\boldsymbol{\beta}$ adalah vektor koefisien model yang berisi γ dan vektor koefisien suku halus individu. Penalti pemulusan total untuk model dapat ditulis sebagai $\sum_j \lambda_j \boldsymbol{\beta}^T \mathbf{S}_j \boldsymbol{\beta}$, dimana λ_j adalah parameter pemulus dan \mathbf{S}_j hanyalah \mathbf{S}_j yang disematkan (*embedded*) sebagai blok diagonal dalam matriks yang mana selain elemen diagonal berisi nol, sehingga $\lambda_j \boldsymbol{\beta}^T \mathbf{S}_j \boldsymbol{\beta}$ adalah penalti untuk f_j . Model aditif secara umum memiliki bentuk: $g(\mu_i) = \mathbf{A}_i \boldsymbol{\gamma} + \sum_j f_j(x_{ji})$, $y_i \sim \text{EF}(\mu_i, \phi)$. Model aditif dapat dipandang sebagai model linier dengan parameter yang berlebih $g(\mu_i) = \mathbf{X}_i \boldsymbol{\beta}$, $y_i \sim \text{EF}(\mu_i, \phi)$ dan parameternya diduga dengan memaksimalkan $l_p(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) - \frac{1}{2\phi} \sum_j \lambda_j \boldsymbol{\beta}^T \mathbf{S}_j \boldsymbol{\beta}$. Parameter λ_j mengontrol pertukaran antara kecocokan dan kehalusan model (Wood, 2017).

1. Pendugaan parameter $\boldsymbol{\beta}$ jika λ diketahui dapat dilakukan melalui *Penalized Iteratively Reweighted Least Squares* (PIRLS) dengan Langkah sebagai berikut (Wood, 2017):
2. Inisialisasi $\hat{\mu}_i = y_i + \delta_i$ dan $\hat{\eta}_i = g(\hat{\mu}_i)$ dimana δ_i biasanya nol, tetapi mungkin konstanta kecil yang memastikan bahwa $\hat{\eta}_i$ adalah berhingga. Ulangi dua langkah berikutnya hingga konvergen.
3. Hitung data semu $z_i = \frac{g'(\hat{\mu}_i)(y_i - \hat{\mu}_i)}{\alpha(\hat{\mu}_i)} + \hat{\eta}_i$, dan bobot iteratif $w_i = \frac{\alpha(\hat{\mu}_i)}{\{g'(\hat{\mu}_i)^2 V(\hat{\mu}_i)\}}$.
4. Temukan $\hat{\boldsymbol{\beta}}$ untuk meminimalkan kuadrat terkecil terboboti. $\|\mathbf{z} - \mathbf{X}\boldsymbol{\beta}\|_W^2 + \sum_j \lambda_j \boldsymbol{\beta}^T \mathbf{S}_j \boldsymbol{\beta}$ dengan $\|\mathbf{z} - \mathbf{X}\boldsymbol{\beta}\|_W^2 = (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})$ dan $V(\mu)$ adalah fungsi varians yang ditentukan oleh sebaran keluarga eksponensial dan $\alpha(\mu_i) = \left[1 + (y_i - \mu_i) \left\{ \frac{V'(\mu_i)}{V(\mu_i)} + \frac{g''(\mu_i)}{g'(\mu_i)} \right\} \right]$ atau dapat menggunakan pendekatan 'Fisher scoring' dimana Hessian dari kemungkinan log diganti dengan nilai harapannya, sesuai dengan pengaturan $\alpha(\mu_i) = 1$.
5. Perbarui $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ dan $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$.

Jika mengikuti hubungan antara GAM dan model campuran linier terampat (*Generalized Linear Mixed Model*-GLMM) dan mengidentifikasi $\lambda_j \mathbf{S}_j / \phi$ dengan matriks pengaruh acak, maka penduga REML yang sesuai untuk parameter skala adalah $\hat{\phi} = \frac{\|\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_W^2}{n - \tau}$ dengan $\tau = \text{tr}\{(\mathbf{X}^T \mathbf{W} \mathbf{X} + \mathbf{S}_\lambda)^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X}\}$, dan $\mathbf{S}_\lambda = \sum_j \lambda_j \mathbf{S}_j$. τ dapat ditafsirkan sebagai derajat kebebasan efektif, dengan $\mathbf{F} = (\mathbf{X}^T \mathbf{W} \mathbf{X} + \mathbf{S}_\lambda)^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X}$ adalah penghitungan bobot. Derajat kebebasan efektif untuk setiap suku mulus diperoleh dengan menjumlahkan nilai F_{ii} yang sesuai dengan koefisien β_i dari suku mulus.

Fungsi mulus dapat dipandang sebagai pengaruh acak dalam model campuran linier terampat yang parameternya dapat diduga menggunakan metode kemungkinan marginal. Kesulitan utama dengan cara kerja ini adalah bahwa fungsi mulus harus diatur dalam bentuk yang sesuai dengan struktur pengaruh acak. Misalnya, untuk setiap smooth dengan parameter smoothing tunggal yang dapat diekspresikan dalam bentuk $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$ dimana $\boldsymbol{\beta}$ diperlakukan sebagai vektor pengaruh tetap (tidak dihukum) dan $\mathbf{b} \sim N(\mathbf{0}, \mathbf{I}\sigma_b^2)$ diperlakukan sebagai vektor pengaruh acak (Wood *et al.*, 2013). Untuk pengaruh acak lebih dari satu, konstruksi matriks dirancang sedemikian rupa sehingga partisi sederhana dari kolom-kolom matriks model smooth dan vektor parameternya, memungkinkan smooths untuk direpresentasikan dalam bentuk $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1 \mathbf{b}_1 + \mathbf{Z}_2 \mathbf{b}_2 + \dots$, dimana $\boldsymbol{\beta}$ merupakan vektor parameter tetap, dan \mathbf{b}_j adalah vektor independen dari pengaruh acak $\mathbf{b}_j \sim N(\mathbf{0}, \mathbf{I}\sigma_{b_j}^2)$.

Model Aditif Geospasial

Model aditif geospasial adalah model aditif dengan menambahkan fungsi pemulus geospasial seperti *thin plate regression splines* (GLMM+TP) dan *Gaussian process* (GLMM+GP). Secara umum, model *additive* memiliki struktur sebagai berikut $g(\mu_i) = \mathbf{X}_i^* \boldsymbol{\theta} + f_1(x_{1i}) + \dots$ dengan $\mu_i = E(Y_i)$ dan Y_i memiliki sebaran yang termasuk ke dalam keluarga eksponensial.

Pertama, GLMM+TP adalah model aditif yang menggabungkan antara model

GLMM dan model *additive* dengan menambahkan fungsi geo-spasial *thin plate regression splines* (TP). Misalkan $\mathbf{S}_i = (s_{i1}, s_{i2})^T$ menunjukkan titik koordinat longitude dan latitude pada lokasi ke- i , $i = 1, \dots, n$. Misalkan \mathbf{Y}_i adalah peubah respon dan $\mathbf{X}_i = (X_{i1}, \dots, X_{i2})^T$ adalah peubah prediktor pada lokasi \mathbf{S}_i . Kita asumsikan kepekatan peluang $(Y|\mathbf{x}, \mathbf{d}, \mathbf{s})$ termasuk anggota sebaran keluarga eksponensial dengan $\mu(\mathbf{x}, \mathbf{d}, \mathbf{s})$ dimodelkan dengan fungsi hubung $g(\cdot)$ dalam bentuk aditif: $g\{\mu(\mathbf{x}, \mathbf{d}, \mathbf{s})\} = \sum_{k=0}^{p-1} \beta_k x_k + d_i u + \alpha(\mathbf{s})$ dengan u adalah pengaruh acak area, dan $\alpha(\cdot)$ adalah fungsi mulus bivariate TP. Maka model GLMM+TP dapat diformulasikan dalam bentuk matriks:

$$g(E(Y)) = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\mathbf{u} + \mathbf{Z}\mathbf{v}, \dots\dots\dots (2)$$

dengan $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{p-1})^T$ adalah parameter yang tidak diketahui, $\mathbf{X} = [\mathbf{X}_1 | \mathbf{X}_2]$; $\mathbf{X}_1 = [1, \mathbf{x}_{ij}^T]_{1 \leq i \leq n}$; $\mathbf{X}_2 = [\mathbf{s}_{ij}^T]_{1 \leq i \leq n}$, $E \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, $\text{Cov} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \sigma_u^2 \mathbf{I}_m & 0 & 0 \\ 0 & \sigma_v^2 \mathbf{I}_{K_s} & 0 \\ 0 & 0 & \sigma_e^2 \mathbf{I}_n \end{bmatrix}$, \mathbf{D} adalah matrik dengan pengaruh acak yang teramati berukuran $n \times m$ dengan m adalah banyaknya sub-populasi. Sedangkan \mathbf{u} adalah vektor pengaruh acak yang tidak diketahui berukuran $m \times 1$, $\mathbf{u} \sim_{\text{iid}} N(\mathbf{0}, \sigma_u^2 \mathbf{I}_m)$. \mathbf{Z} adalah matriks berukuran $n \times K_s$ dari bivariate TP berdasarkan fungsi: $\mathbf{Z} = [C(s_i - \kappa_k^s)]_{1 \leq i \leq n, 1 \leq k \leq K_s}$, dan $C(\mathbf{s}) = \|\mathbf{s}\|^2 \log \|\mathbf{s}\|$ dan $\kappa_1^s, \dots, \kappa_{K_s}^s$ adalah simpul (knots). \mathbf{v} adalah vektor koefisien bivariate TP berukuran $K_s \times 1$, $\mathbf{v} \sim_{\text{iid}} N(\mathbf{0}, \sigma_v^2 \mathbf{I}_{K_s})$. Sedangkan \mathbf{e} adalah galat acak, $\mathbf{e} \sim_{\text{iid}} N(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$.

Kedua, GLMM+GP adalah model aditif yang menggabungkan antara model GLMM dan model *additive* dengan menambahkan fungsi geo-spasial *Gaussian Process* (GP). Misalkan $\mathbf{S}_i = (s_{i1}, s_{i2})^T$ menunjukkan titik koordinat longitude dan latitude pada lokasi ke- i , $i = 1, \dots, n$. Misalkan \mathbf{Y}_i adalah peubah respon dan $\mathbf{X}_i = (X_{i1}, \dots, X_{i2})^T$ adalah peubah prediktor pada lokasi \mathbf{S}_i . Kita asumsikan kepekatan peluang $(Y|\mathbf{x}, \mathbf{d}, \mathbf{s})$ termasuk

anggota sebaran keluarga eksponensial dengan $\mu(\mathbf{x}, \mathbf{d}, \mathbf{s})$ dimodelkan dengan fungsi hubung $g(\cdot)$ dalam bentuk aditif. Jika $\text{var}(Y|\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}) = \sigma^2 V\{\mu(\mathbf{x}, \mathbf{s})\}$, maka pendugaan nilai tengah (*mean*) dapat diperoleh dengan mengganti fungsi log-likelihood bersyarat ($\log\{f_{Y|\mathbf{x}, \mathbf{s}}(y|\mathbf{x}, \mathbf{s})\}$) dengan fungsi quasi-likelihood $l(\vartheta, y)$, yang memenuhi $\nabla_{\vartheta} l(\vartheta, y) = \frac{y - \vartheta}{\sigma^2 V(\vartheta)}$. Pendugaan ini menggunakan pendekatan quasi-likelihood nonparametrik (Yu *et al* 2019).

Misalkan $\alpha(\cdot)$ adalah fungsi *additive Gaussian Process*, maka fungsi α adalah penjumlahan dari k fungsi regresi yang dikontrol oleh himpunan parameter $\boldsymbol{\phi} = (\phi_1, \dots, \phi_k)^T$ atau $\alpha(s_i) = \phi_1 f_1(s_1) + \dots + \phi_k f_k(s_k)$. Maka GLMM+GP dapat diformulasikan mirip dengan GLMM+TP dalam bentuk matriks $g(E(Y)) = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\mathbf{u} + \mathbf{P}\boldsymbol{\theta}$. *Gaussian process* (GP) menyediakan prior yang fleksibel untuk setiap fungsi komponen dalam $\{f_l, l = 1, \dots, k\}$ dimana $f_l \sim GP(0, c_l)$ dengan $c_l(s, s') = \exp\left\{-\sum_{j=1}^p K_{ij}(s_j - s'_j)^2\right\}$ dengan hyperprior Gamma yang ditetapkan ke parameter inverse-bandwidth K_{ij} memastikan pendugaan optimal dari fungsi regresi isotropic. Pendugaan parameter dari model ini dapat melihat Vo & Pati (2017).

Metode Evaluasi

Gugus data dibagi menjadi data training dan data testing dengan perbandingan 70:30 persen. Rasio perbandingan antara data training dan testing ditentukan oleh peneliti dengan asumsi terjadi non-respon sebanyak 30 persen. Model dibangun berdasarkan data training yang masing-masing gugus data diulang sebanyak 1000 kali. Kemudian dilakukan pendugaan berat gabah kg per $2.5 \times 2.5 m^2$ menggunakan ketiga model tersebut. *Mean Squared Error of Prediction* (MSEP) dihitung dengan rumus:

$$MSEP = \frac{\sum_{i=1}^{n_{data\ testing}} [y_i - \hat{y}_i]^2}{n_{data\ testing}}, \dots\dots\dots (3)$$

dimana y_i adalah data aktual (*true value*) dan \hat{y}_i adalah data dugaan. Model terbaik dipilih berdasarkan rata-rata MSEP terkecil dari 1000 ulangan.

Setelah diperoleh model terbaik, dilakukan perbandingan rata-rata produktivitas padi antara teknik menghapus data dan teknik imputasi data non-respons. Teknik yang memberikan hasil lebih mirip dengan rata-rata pada data lengkap maka teknik tersebut lebih baik.

Evaluasi dilakukan dengan membandingkan tingkat presisi metode imputasi melalui tiga skenario data non-respons. Data lengkap dibuat seolah-olah terdapat data non-respons sebanyak 30 persen. Teknik menghapus data hilang artinya rata-rata produktivitas dihitung dari 70 persen data. Teknik imputasi artinya menduga 30 persen data hilang dengan model yang diajukan kemudian digabung dengan 70 persen data respon. Kemudian digabung dan dihitung rata-

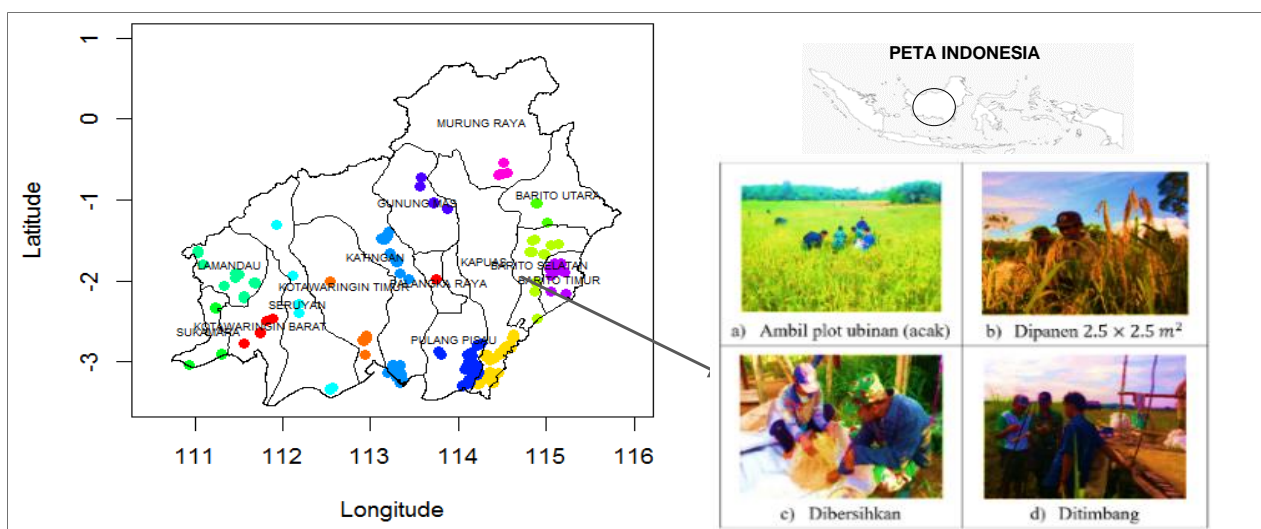
rata produktivitas padi. Data 30 persen non-respons diatur dengan skenario tiga kasus, yaitu (1) data hilang terjadi pada kelompok dengan produktivitas padi tinggi, (2) data hilang menyebar secara acak, dan (3) data hilang terjadi pada kelompok dengan produktivitas padi rendah.

Data dan Sumber Data

Data yang digunakan untuk pemodelan adalah data hasil Survei Ubinan di Kalimantan Tengah Tahun 2019. Data dikumpulkan oleh petugas dari Badan Pusat Statistik dibantu penyuluh lapangan dari Dinas Pertanian. Data dikumpulkan saat petani sedang memanen tanaman padi pada petak terpilih. Peubah penelitian dapat dilihat pada Tabel 1.

Tabel 1. Peubah yang digunakan untuk pemodelan

Peubah	Nama Peubah	Sumber berasal dari Survei Ubinan 2019	Keterangan
y	Berat gabah kering panen	Rincian 701	Peubah respon
x_1	Varietas benih (Hibrida atau inbrida)	R609	
x_2	Banyaknya pupuk urea (kg/ha)	R610_1 per R604 kali 10000	
x_3	Banyaknya TSP/SP36 (kg/ha)	R610_2 per R604 kali 10000	
x_4	Banyaknya pupuk KCL (kg/ha)	R610_3 per R604 kali 10000	
x_5	Banyaknya pupuk NPK (kg/ha)	R610_4 per R604 kali 10000	
x_6	Banyaknya pupuk kompos (kg/ha)	R610_5 per R604 kali 10000	
x_7	Serangan OPT (terserang atau tidak)	R804b	Pengaruh tetap
x_8	Dampak perubahan iklim (terdampak atau tidak)	R805b	
x_9	Kabupaten	R102	Pengaruh acak ^{*)}
x_{10}	Titik Longitude	R303	Pengaruh aditif geospasial
x_{11}	Titik Latitude	R303	



Gambar 1. Sebaran titik sampel Survei Ubinan di Kalimantan Tengah, Tahun 2019

Data lengkap yang terkumpul pada pelaksanaan Survei Ubinan tanaman padi di Kalimantan Tengah Tahun 2019 adalah sebanyak 585 titik plot yang tersebar di 14 kabupaten/kota se-Kalimantan Tengah. Sebaran titik contoh Survei Ubinan di Kalimantan Tengah tahun 2019 dapat dilihat pada Gambar 1.

HASIL DAN PEMBAHASAN

Beberapa titik sampel Survei Ubinan di Kalimantan Tengah terletak pada lokasi yang berdekatan. Untuk menguji adanya ketergantungan spasial maka digunakan Uji Indeks Moran. Nilai indeks moran diperoleh sebesar $I = 0.4423$ dengan p-value 0.000 sehingga cukup bukti untuk mengatakan terjadi ketergantungan spasial pada data produktivitas padi. Semakin dekat lokasi penanaman padi maka akan semakin mirip nilai produktivitas padi yang dihasilkan. Hal ini menjadi argumen dalam memasukkan fungsi pemulus geospasial ke dalam model. Hasil pendugaan parameter pengaruh tetap dapat dilihat pada Tabel 2.

Tabel 2 menunjukkan bahwa hasil dugaan parameter pengaruh tetap pada ketiga model adalah mirip. Faktor yang berpengaruh nyata terhadap produktivitas padi di Kalimantan Tengah adalah x_1 (Varietas benih), x_2 (Pupuk urea), x_3 (Pupuk

TSP/SP36), x_5 (Pupuk NPK), x_7 (Terkena serangan OPT atau tidak), x_8 (terkena dampak perubahan iklim atau tidak). Sedangkan x_4 (Pupuk KCL) dan x_6 (Pupuk kompos) tidak berpengaruh nyata terhadap produktivitas padi di Kalimantan Tengah. Dugaan ragam pengaruh acak dari ketiga model dapat dilihat pada Tabel 3.

Terlihat pada Tabel 3 bahwa saat dilakukan penambahan fungsi pemulus maka dugaan ragam pengaruh acak kabupaten menjadi menyusut. Nilai Dugaan ragam fungsi pemulus $\widehat{\sigma}_v^2$ lebih besar dibanding dugaan ragam pengaruh acak kabupaten $\widehat{\sigma}_u^2$. Artinya, penambahan fungsi pemulus dapat meningkatkan kinerja prediksi model. Perbandingan nilai MSEP antarmodel disajikan pada Gambar 2.

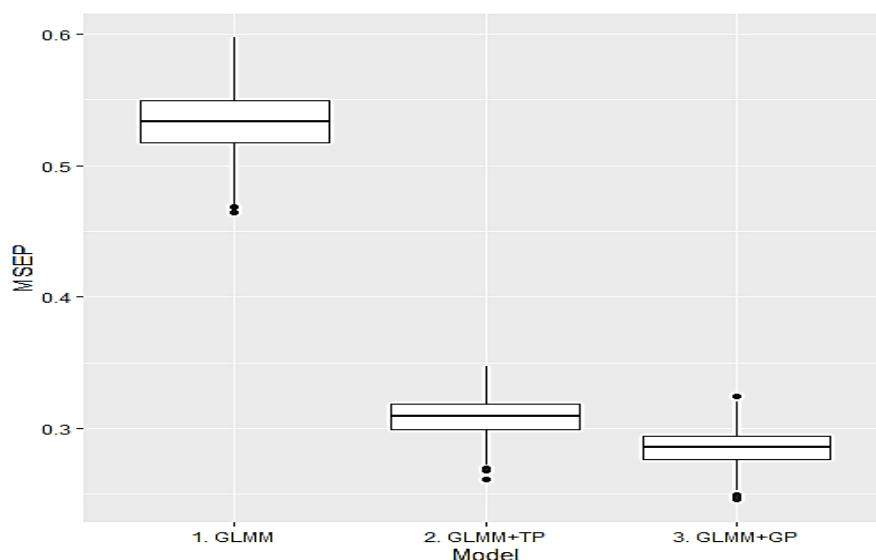
Dapat dilihat dari Gambar 2 bahwa GLMM+GP memiliki nilai MSEP terkecil disusul GLMM+TP dan GLMM. Nilai Quartil 1 (Q1) pada MSEP GLMM+TP lebih besar dibanding nilai Q3 pada MSEP GLMM+GP. Hal ini mengindikasikan bahwa model GLMM+GP memiliki performa terbaik secara nyata. Dengan pengulangan 1000 kali maka derajat bebas galat sebesar $1000 - 1 = 999$ yang menyebabkan kuadrat tengah galat menjadi sangat kecil mendekati nol. Oleh karena itu, dapat ditunjukkan bahwa model GLMM+GP merupakan model yang

Tabel 2. Nilai dugaan parameter pada model GLMM, GLMM+TP, dan GLMM+GP

Peubah	Dugaan parameter			Pr(> t)		
	GLMM	GLMM+TP	GLMM+GP	GLMM	GLMM+TP	GLMM+GP
Intersep	2.2800	2.4470	2.4402	0.0000	0.0000	0.0000
x_1	-0.7901	-0.5452	-0.5144	0.0000	0.0000	0.0000
x_2	0.0015	0.0016	0.0014	0.0034	0.0004	0.0016
x_3	0.0027	0.0022	0.0020	0.0000	0.0006	0.0023
x_4	-0.0010	-0.0006	-0.0011	0.6850	0.7680	0.6014
x_5	0.0024	0.0014	0.0013	0.0000	0.0000	0.0002
x_6	0.0005	0.0007	0.0008	0.2808	0.0980	0.0796
x_7	0.2680	0.1331	0.1413	0.0000	0.0246	0.0175
x_8	0.1574	0.2076	0.2004	0.0259	0.0011	0.0018

Tabel 3. Dugaan ragam pengaruh acak pada GLMM, GLMM+TP, GLMM+GP

Pengaruh Acak	GLMM	GLMM+TP	GLMM+GP
$\widehat{\sigma}_u^2$	0.2017	0.0006	0.0064
$\widehat{\sigma}_v^2$	-	10.5924	2.6835
$\widehat{\sigma}_e^2$	0.5515	0.5904	0.5740



Gambar 2. Perbandingan MSEP antara GLMM, GLMM+TP, dan GLMM+GP

memberikan performa terbaik secara nyata dalam pendugaan produktivitas padi dibanding dua model lainnya. Model GLMM+GP dapat digunakan untuk mengimputasi data hilang berat gabah pada Survei Ubinan pada pelaksanaan Survei Ubinan berikutnya.

Kajian data non-respons

Pada bagian ini digunakan data asli hasil Survei Ubinan 2019 di tiga kabupaten penghasil padi terbesar di Kalimantan Tengah (Kapuas, Pulang Pisau, dan katingan). Langkah yang diterapkan pada kajian ini adalah sebagai berikut: Pertama, data ubinan yang lengkap dibagi menjadi 75 persen data respon dan 25 persen data non-respon.

Pengambilan 25 persen data non-respon dilakukan secara acak dengan tiga skenario, yaitu data hilang terjadi pada kelompok dengan produktivitas padi tinggi (kasus I), data hilang menyebar secara acak (kasus II), dan data hilang terjadi pada kelompok dengan produktivitas padi yang rendah (kasus III). Langkah kedua, pemodelan dilakukan menggunakan data lengkap. Ketiga, menduga produktivitas padi menggunakan model pada tahap dua. Selanjutnya, mengulangi ketiga langkah tersebut sebanyak 200 iterasi.

Kajian ini menunjukkan bahwa teknik imputasi lebih baik dibanding teknik menghapus data non-respons. Teknik mengabaikan non-respon (*listwise deletion*)

Tabel 4. Perbandingan rata-rata berat gabah seluas $2.5 \times 2.5 m^2$ (\bar{y}) antara teknik mengabaikan data non-respons, teknik imputasi dengan GLMM+GP, dan data lengkap

Jenis data hilang	\bar{y} pada data lengkap	\bar{y} dengan mengabaikan non-respons		\bar{y} setelah ditambah hasil imputasi	
		dugaan	selisih (2) - (3)	dugaan	selisih (2) - (5)
(1)	(2)	(3)	(4)	(5)	(6)
Kasus I ¹⁾					
Katingan	1.96	1.49	0.47	1.61	0.35
Pulang Pisau	2.23	1.74	0.49	1.85	0.38
Kapuas	2.58	2.15	0.43	2.24	0.34
Kasus II ²⁾					
Katingan	1.96	2.01	-0.05	1.94	0.02
Pulang Pisau	2.23	2.19	0.04	2.21	0.02
Kapuas	2.58	2.54	0.04	2.54	0.04
Kasus III ³⁾					
Katingan	1.96	2.39	-0.43	2.22	-0.26
Pulang Pisau	2.23	2.68	-0.45	2.54	-0.31
Kapuas	2.58	2.95	-0.37	2.79	-0.21

hanya bisa diterapkan dalam kasus data hilang tersebar secara acak atau *Missing at Random* (MAR). Dalam praktiknya, pada pelaksanaan Survei Ubinan di Kalimantan Tengah, seringkali non-respons terjadi tidak secara acak. Oleh karena itu, metode imputasi disarankan untuk diterapkan dalam mengatasi missing data yang disebabkan non-respons.

Dapat dilihat pada Tabel 4 bahwa saat data hilang tersebar secara acak (kasus II) maka angka rata-rata berat gabah mirip dengan rata-rata saat tidak ada data hilang. Artinya, saat data hilang tersebar acak maka metode mengabaikannya dapat diterapkan. Berbeda dengan jenis data hilang pada kasus I dan III, rata-rata produktivitas yang dihasilkan bisa *underestimate* ataupun *overestimate* sehingga perlu ditangani. Terlihat dalam kasus I dan III atau kasus MNAR (*Missing Not at Random*), teknik imputasi data hilang menghasilkan rata-rata yang lebih mirip dengan rata-rata data lengkap. Dapat ditunjukkan bahwa metode imputasi bermanfaat saat kondisi missing values adalah MNAR karena mampu mempersempit bias dugaan.

KESIMPULAN

Terdapat tiga faktor utama yang berpengaruh nyata terhadap produktivitas padi di Kalimantan Tengah tahun 2019 yaitu varietas, pupuk, dan faktor lingkungan. Pupuk yang berpengaruh nyata adalah pupuk urea, TSP/SP36, dan NPK/ pupuk majemuk, sedangkan faktor lingkungan yang berpengaruh nyata adalah serangan OPT dan dampak perubahan iklim. Berdasarkan indeks Moran, terjadi ketergantungan spasial pada data produktivitas padi di Kalimantan Tengah. Semakin dekat lokasi penanaman padi maka akan semakin mirip nilai produktivitas padinya. Hal ini menjadi argumen dalam memasukkan fungsi pemulus geospasial ke dalam model. Hasil perbandingan kinerja antarmodel menunjukkan bahwa model dengan penambahan fungsi pemulus *Gaussian Process* memberikan kinerja terbaik dengan nilai MSEP tekecil secara signifikan. Berdasarkan kajian, dapat ditunjukkan bahwa metode imputasi bermanfaat saat data non-respons menyebar tidak acak atau kondisi missing valuesnya adalah MNAR. Sedangkan

pada kasus MAR teknik *listwise deletion* sama baiknya dibanding teknik imputasi. Berdasarkan hasil penelitian ini, saat gagal melakukan pengukuran berat gabah, petugas BPS disarankan tetap melakukan wawancara kepada petani untuk memperoleh peubah-peubah yang dapat dimanfaatkan untuk mengimputasi data hilang.

DAFTAR PUSTAKA

- Ardiansyah, M., & Tofri, Y. (2019). Perbandingan Data Produktivitas Padi Antara Hasil Wawancara Pascapanen dengan Data Survei Ubinan di Kalimantan Tengah. *Jurnal Penelitian Pertanian Tanaman Pangan*, 3(1), 17-22. doi:<http://dx.doi.org/10.21082/jpntp.v3n1>.
- Ardiansyah, M., Buana, W. P., & Kurnia, A. (2020). Prediksi Produktivitas Padi Melalui Survei Ubinan Menggunakan Model Linier dan Quantile Regression Forest. *Jurnal Penelitian Pertanian Tanaman Pangan*, 4(3), 135-144. doi:<http://dx.doi.org/10.21082/jpntp.v4n3>.
- Ardiansyah, M., Kurnia, A., Sadik, K., Djuraidah, A., & Wijayanto, H. (2021). Numerical Prediction of paddy weight of Crop Cutting Survey using Generalized Geoadditive Linear Mixed Model. *Journal of Physics: Conference Series*, 1863 (2021) 012024, 1-17. doi:10.1088/1742-6596/1863/1/012024
- Breslow, N. E., & Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88(421), 9-25.
- Chhabra, G., Vashisht, V., & Ranjan, J. (2019). A Review on Missing Data Value Estimation Using Imputation Algorithm. *Journal of Advanced Research in Dynamical and Control Systems*, 11(07), 312-318.
- Curley, C., Krause, R. M., Feiock, R., & Hawkins, C. V. (2019). Dealing with Missing Data: A Comparative Exploration of Approaches Using the Integrated City Sustainability Database.

- Urban Affairs Review*, 55(2), 591-615.
doi:10.1177/1078087417726394.
- Djuraidah, A. (2020). *Monograf Penerapan dan Pengembangan Regresi Spasial dengan Studi Kasus pada Kesehatan, Sosial, dan Ekonomi*. Bogor: IPB Press.
- Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, 1(3), 297-318.
- Lin, X., & Zhang, D. (1999). 1999. Inference in generalized additive mixed models by using smoothing splines. *Royal Statistical Society*, 61(2), 381-400.
- Nelder, J. A., & Wedderburn, R. W. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society*, 135(3), 370-384.
- Vo, G., & Pati, D. (2017). Sparse Additive Gaussian Process with Soft Interactions. *Open Journal of Statistics*, 7, 567-588.
doi:10.4236/ojs.2017.74039.
- Wood, S. N. (2017). *Generalized Additive Models An Introduction with R Second Edition*. London: Chapman & Hall/CRC Press, Taylor & Francis Group.
- Wood, S. N., Scheipl, F., & Faraway, J. J. (2013). Straightforward intermediate rank tensor product smoothing in mixed models. *Stat Comput*, 23, 341-360.
doi:10.1007/s11222-012-9314-z.
- Yu, S., Wang, G., Wang, L., Liu, C., & Yang, L. (2019). Estimation and Inference for Generalized Geoadditive Models. *Journal of the American Statistical Association*, 1-14.
doi:10.1080/01621459.2019.1574584.

METODE HUTAN ORDINAL UNTUK KLASIFIKASI DESA SESUAI DENGAN STATUS INDEKS DESA MEMBANGUN

Dwi Agustin Nuriani Sirodj^{1,2}, Khairil Anwar Notodiputro¹, Bagus Sartono¹

¹Departemen Statistika, FMIPA, IPB University, Indonesia

²Program Studi Statistika, FMIPA, Universitas Islam Bandung, Indonesia

e-mail: bagusco@apps.ipb.ac.id

Received: May 18, 2023; Revised: October 4, 2023; Accepted: December 29, 2023; Published: December 31, 2023

Abstrak

Metode hutan acak merupakan metode klasifikasi berbasis pohon yang cukup populer. Metode ini jika diterapkan pada peubah respons ordinal akan memberikan hasil yang mengubah karakteristik dari data ordinal itu sendiri. Dalam tulisan ini akan dibahas kinerja dari metode hutan ordinal dan hutan ordinal Naïve untuk klasifikasi status berbagai desa sesuai dengan Indeks Desa Membangun (IDM) di kabupaten Tasikmalaya dan Indramayu. Hasilnya memperlihatkan bahwa untuk Kabupaten Tasikmalaya kinerja metode hutan ordinal mengungguli kinerja hutan ordinal Naïve dengan rata-rata akurasi sebesar 73.8% dan rata-rata nilai kappa sebesar 0.18 sedangkan pada Kabupaten Indramayu kinerja metode hutan ordinal Naïve yang justru mengungguli metode hutan ordinal dengan rata-rata akurasi sebesar 51.6% dan rata-rata nilai kappa sebesar 0.162. Selain itu ditunjukkan pula bahwa peubah yang penting dalam proses klasifikasi status IDM di Kabupaten Tasikmalaya dan Kabupaten Indramayu adalah peubah Jumlah Koperasi dan Jarak terdekat ke rumah sakit.

Kata kunci: Hutan ordinal, Naïve Hutan ordinal, Klasifikasi Status IDM

Abstract

The random forest (RF) method is a popular tree-based classification method. However, when faced with ordinal response variables, the prediction results of the RF method cannot maintain the characteristics of the ordinal data itself. So, in this paper discusses the performance of the ordinal forest and Naïve ordinal forest methods for the rural's classification based on the Village Development Index (IDM) in Tasikmalaya and Indramayu. The results show that in the Tasikmalaya district, the OF method performs better, with an average accuracy of 73.8% and an average kappa value of 0.18. However, for the Indramayu district, the Naïve OF method looks better than the OF method, producing an average accuracy of 51.6% and an average kappa value of 0.162. In addition, it is also shown that the important variables in the rural's classification in Tasikmalaya and Indramayu districts are the variables of the number of cooperatives and the closest distance to the hospital.

Keywords: Ordinal Forest, Naïve Ordinal Forest, rural's classification

PENDAHULUAN

Metode hutan acak (RF) pertama kali diperkenalkan oleh (Breiman, 2001) banyak diaplikasikan pada saat ingin melakukan klasifikasi berbasis pohon pada data berdimensi tinggi. RF didasarkan pada teknik pohon keputusan sehingga mampu mengatasi masalah non-linier (Dewi et al., 2011). RF termasuk kedalam metode statistik non parametrik yang juga dapat diterapkan pada kasus peubah respons yang berkorelasi, adanya interaksi pada peubah bebas dan juga data dengan tingkat heterogenan yang tinggi. (Hothorn et al., 2006) mengenalkan hutan acak yang berbasis inferensial bersyarat, yang menyediakan pemilihan variabel yang tidak bias ketika mencari *split* yang optimal. Pada saat dihadapkan peubah respons nominal atau numerik, RF mampu memberikan kinerja yang baik, hanya saja ketika terdapat peubah respons ordinal, tidak ada prosedur yang standar sehingga hasil prediksi tidak dapat mempertahankan karakteristik dari data ordinal itu sendiri (Janitza et al., 2016).

Pengembangan dilakukan oleh (Hornung, 2017) yang memperkenalkan RF pada data ordinal atau lebih dikenal dengan metode hutan ordinal (OF). Pada metode OF peubah respons ordinal diubah ke dalam nilai skor yang kemudian akan diperlakukan sebagai peubah kontinu/ data numerik. Prinsip yang dilakukan pada OF adalah mencari nilai skor yang optimal yang akan menggantikan pelabelan pada peubah respons ordinal kemudian mengonstruksi OF seperti pohon regresi dalam RF. Dalam proses mencari nilai skor pada setiap kategori, (Hornung, 2020) memberikan alternatif dengan tetap menggunakan pelabelan awal pada peubah respons ordinal, atau yang dikenal sebagai metode Naïve OF. Metode OF juga dapat digunakan untuk memeringkat kovariat berdasarkan kepentingannya dalam prediksi. kemungkinan untuk memperhitungkan informasi tingkatan pada peubah respons. Identifikasi peubah penjas yang berhubungan dengan peubah respons pada

OF dapat ditentukan melalui *Variable Importance Measures* (VIMs). Menurut (Janitza et al., 2016) terdapat tiga jenis ukuran VIMs yang dapat digunakan pada respons ordinal, yaitu *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), dan *Ranked Probability Score* (RPS) yang ketiganya mempertimbangkan informasi tingkatan pada peubah respons.

Beberapa penelitian terkait OF diantaranya (Irawan, 2019) yang mengkaji prevalensi kerawanan pangan dan peubah-peubah rumah tangga (RT) menggunakan metode OF dan Regresi Logistik Ordinal (RLO). Dalam hal klasifikasi RT ke dalam pengategorian rawan pangan, OF mencatat kinerja lebih baik daripada RLO. (Nisa, 2019) mengidentifikasi faktor-faktor yang memengaruhi prestasi mahasiswa FMIPA IPB angkatan 2015-2017 menggunakan RLO dan OF. Model terbaik yang didapatkan dari penelitian ini adalah model RLO. Selanjutnya (Agustine, 2019) menggunakan metode OF dalam menentukan peubah pencari terbaik dalam pengklasifikasian hasil akreditasi Sekolah Menengah Atas Negeri. Selanjutnya (Hornung, 2020) membandingkan kinerja dari metode OF, Naïve OF, *multi-class* RF, dan *ordered probit regression* dalam klasifikasi peubah respons ordinal. Secara umum metode OF memiliki kinerja yang baik di bandingkan dengan metode lainnya dalam memprediksi peubah respons ordinal.

Disparitas kemiskinan perkotaan dan perdesaan di Provinsi Jawa Barat masih tergolong tinggi. Hal ini dapat terlihat dari laporan Badan Pusat Statistik (BPS) Provinsi Jawa Barat mengenai perkembangan tingkat kemiskinan pada maret 2022 bahwa selama periode September 2021-Maret 2022, jumlah penduduk miskin perkotaan bertambah 59,29 ribu orang dan di perdesaan naik 6,84 ribu orang. Persentase kemiskinan di perkotaan meningkat dari 7,48 persen menjadi 7,57 persen. Sementara itu, di perdesaan juga mengalami kenaikan dari 9,76 persen menjadi 9,88 persen pada periode yang sama (*Badan Pusat Statistik Provinsi Jawa Barat*, n.d.). Kabupaten

Indramayu dan Kabupaten Tasikmalaya masih termasuk kedalam 10 daerah di Propinsi Jawa Barat dengan rata-rata persentase penduduk miskin tertinggi pada tahun 2021. Kabupaten Indramayu dengan rata-rata persentase penduduk miskin sebesar 13.82% dan kabupaten tasikmalaya dengan rata-rata persentase penduduk miskin sebesar 11.04%.

Desa sebagai satuan pemerintahan terkecil mempunyai peranan penting dalam mendukung pembangunan berkelanjutan di Indonesia. Desa yang juga berperan sebagai subjek pelaku pembangunan pada level terendah menjadi ujung tombak terkait maju atau tidaknya suatu wilayah. Dalam Permendagri No.114 Tahun 2014, pembangunan desa adalah upaya peningkatan kualitas hidup dan kehidupan untuk sebesar-besarnya kesejahteraan masyarakat desa.

Indeks Desa Membangun (IDM) adalah Indeks Komposit yang dibentuk dari: (1) Indeks Ketahanan Sosial; (2) Indeks Ketahanan Ekonomi; dan Indeks Ketahanan Ekologi. IDM menjadi dasar penetapan status kemajuan dan kemandirian desa; dan juga penyedia data dan informasi dasar bagi pembangunan Desa. IDM juga berfungsi sebagai peta pengembangan pembangunan yang ada di Desa (Setyowati, 2019). IDM mengklasifikasi desa dalam lima kategori yaitu (i) Desa Sangat Tertinggal; (ii) Desa Tertinggal; (iii) Desa Berkembang; (iv) Desa Maju; dan (v) Desa Mandiri (Hanibal, 2015). Klasifikasi desa tersebut untuk menunjukkan keragaman karakter setiap desa serta digunakan pula untuk menajamkan penetapan status perkembangan Desa dan sekaligus rekomendasi intervensi kebijakan yang diperlukan. Beberapa penelitian terdahulu terkait IDM diantaranya pengelompokan desa berdasarkan status IDM menggunakan metode *Gaussian Mixture Model* menghasilkan tujuh kluster dengan nilai *Bayesian Information Criteria* sebesar 9979,9 (Narulita & Kudus, 2022), dan pemodelan klasifikasi status IDM menggunakan algoritma C5.0 (Fitria et al., 2022) dengan nilai akurasi sebesar 82,35%.

Dari dua metode sebelumnya kedudukan peubah respons status IDM yang mempunyai skala ordinal dianalisis seperti peubah kategori pada umumnya, akan tetapi pada metode OF dan Naïve OF akan tetap dipertahankan kedudukannya sebagai peubah respons berskala ordinal.

Adapun tujuan dari penelitian ini adalah untuk melakukan kajian terhadap kinerja dari metode OF dan Naïve OF dalam proses klasifikasi status IDM di Kabupaten Tasikmalaya dan Kabupaten Indramayu berdasarkan faktor sosial, ekonomi, ekologi dan juga profil dari perangkat desa serta mengidentifikasi faktor-faktor yang memengaruhi klasifikasi status IDM di Kabupaten Tasikmalaya dan kabupaten Indramayu di tahun 2021.

METODOLOGI

Tinjauan Referensi

Dalam OF, peubah respons ordinal diperlakukan sebagai peubah kontinu. Prinsipnya adalah melakukan perubahan dari yang semula berupa data kategori ordinal ke dalam nilai skor yang kemudian nilai skor ini yang akan diperlakukan sebagai peubah kontinu. OF pada dasarnya melakukan *regression forest* pada nilai skor $s_1 < s_2 < \dots < s_j$ yang akan menggantikan nilai kelas $1, 2, \dots, j$ pada peubah respons ordinal. Nilai skor $s_1 < s_2 < \dots < s_j$ dioptimalkan sedemikian rupa sehingga kinerja klasifikasinya menjadi optimal. Sehingga terdapat dua hal yang akan dilakukan pada OF adalah (1) mengoptimalkan nilai skor yang terbentuk, dan (2) mengonstruksi *regression forest* dari nilai skor yang telah optimal.

1. Algoritma Hutan ordinal

Misalkan suatu contoh $\{(x_1, y_1), \dots, (x_n, y_n)\}$, dengan x_i ($i \in \{1, \dots, n\}$) merupakan vektor kovariat-kovariat amatan i dan $y_i \in \{1, \dots, J\}$ merupakan nilai kelas peubah respons ordinal untuk amatan tersebut. Maka aturan pembentukan prediksi OF adalah sebagai berikut (Hornung, 2020):

1. Untuk $b = 1, \dots, Bsets$ (contohnya, $Bsets = 1.000$):

- a) Tariklah sebanyak $J - 1$ nilai dari peubah acak bersebaran seragam $U(0,1)$ lalu urutkan nilai-nilai tersebut. Nilai yang terurut disimbolkan dengan $d_{b,2}, \dots, d_{b,J}$. Untuk melengkapinya, definisikan $d_{b,1} := 0$ dan $d_{b,J+1} := 1$.
 - b) Bentuklah peubah respons kontinu $\mathbf{z}_b = z_{b,1}, \dots, z_{b,n}$ dengan mengganti setiap nilai kelas $j, j = 1, \dots, J$, pada peubah respons ordinal $\mathbf{y} = y_1, \dots, y_n$ dengan nilai ke- j di himpunan skor $\mathbf{s}_b := \{s_{b,1}, \dots, s_{b,J}\}$, dengan $s_{b,j} := \Phi^{-1}(c_{b,j})$ dan $c_{b,j} := (d_{b,j} + d_{b,j+1})/2$ ($j \in \{1, \dots, J\}$).
 - c) Buat regression forest f_{sb} dengan $Bntreeprior$ pohon (contohnya, $Bntreeprior = 100$) menggunakan \mathbf{z}_b sebagai peubah respons.
 - d) Dapatkan prediksi OBB $\hat{z}_{b,1}, \dots, \hat{z}_{b,n}$ dari $z_{b,1}, \dots, z_{b,n}$.
 - e) Dapatkan prediksi OBB dari y_1, \dots, y_n dengan cara $\hat{y}_{b,i} = j$ jika $\hat{z}_{b,i} \in [\Phi^{-1}(d_{b,j}), \Phi^{-1}(d_{b,j+1})]$ ($i \in \{1, \dots, n\}$).
 - f) Berikan skor kinerja $sc_b := g(\mathbf{y}, \hat{\mathbf{y}}_b)$ kepada f_{sb} , dengan $\hat{\mathbf{y}}_b := \hat{y}_{b,1}, \dots, \hat{y}_{b,n}$ dan g merupakan fungsi khusus (diistilahkan dengan fungsi kinerja) yang pemilihannya tergantung konteks penelitian.
2. Misalkan S_{best} adalah himpunan indeks-indeks dari $B_{bestsets}$ (contohnya, $B_{bestsets} = 10$) regression forest yang terbentuk pada poin 1 dengan nilai sc_b terbesar. Maka untuk setiap $j \in \{1, \dots, J + 1\}$ ambillah rata-rata dari nilai-nilai $d_{b,j}$ tersebut dengan ketentuan $b \in S_{best}$, sehingga dihasilkan suatu himpunan nilai $J + 1$ yang dinyatakan sebagai d_1, \dots, d_{J+1} .
 3. Bentuklah peubah respons kontinu yang baru $\mathbf{z} = z_1, \dots, z_n$ dengan mengganti setiap nilai kelas $j, j = 1, \dots, J$ dalam peubah respons ordinal $\mathbf{y} = y_1, \dots, y_n$ dengan nilai ke- j dalam himpunan skor optimal $\{s_1, \dots, s_J\}$, dengan $s_j = \Phi^{-1}(c_j)$ dan $c_j = (d_j + d_{j+1})/2$ ($j \in \{1, \dots, J\}$).
 4. Buat regression forest f_{final} dengan $Bntree$ pohon (contohnya, $Bntree = 5.000$) menggunakan \mathbf{z} sebagai peubah respons.

2. Prediksi dengan OF

Prediksi nilai peubah respons dari suatu amatan independen i^* dengan vektor kovariatnya \mathbf{x}_{i^*} didapat dengan cara sebagai berikut (Hornung, 2020) :

1. Untuk $b = 1, \dots, B_{ntree}$:
 - a. Terapkan pohon ke- b dalam f_{final} ke amatan i^* dan dapatkan prediksi $\hat{z}_{i^*,b}$
 - b. Dapatkan prediksi kelas dari pohon ke- b : $\hat{y}_{i^*,b} = j$ jika $\hat{z}_{i^*,b} \in [\Phi^{-1}(d_{b,j}), \Phi^{-1}(d_{b,j+1})]$.
2. Dapatkan prediksi kelas final dari $\hat{y}_{i^*,1}, \dots, \hat{y}_{i^*,B_{ntree}}$ dengan teknik suara terbanyak.

3. Akurasi Model

Evaluasi yang digunakan dalam penelitian ini adalah nilai akurasi. Evaluasi digunakan untuk melihat tingkat kesalahan yang terjadi pada klasifikasi area contoh sehingga dapat dilihat besar persentase ketelitiannya. Akurasi adalah keakuratan kinerja klasifikasi atau persentase objek yang diklasifikasikan dengan benar terhadap keseluruhan data uji.

Tabel 1. Confusion Matrix

Klasifikasi dari Data Latih	Klasifikasi dari Data Uji			Total
	$j = 1$	\dots	$j = J$	
$j = 1$	n_{11}	\dots	n_{1J}	$n_{1.}$
\dots	\dots	\dots	\dots	\dots
$j = J$	n_{J1}	\dots	n_{JJ}	$n_{J.}$
Total	$n_{.1}$	\dots	$n_{.j}$	$n_{..}$

$$TTP_{all} = \sum_{j=1}^J n_{jj} \quad [1]$$

$$akurasi = \frac{TTP_{all}}{n_{..}} \quad [2]$$

Dimana TTP_{all} : Total banyaknya amatan kelas i yang benar diklasifikasikan sebagai kelas j .

Selain melihat nilai akurasi, ukuran lain yang digunakan adalah *unweighted Kappa statistic* dengan rumus sebagai berikut (Sim & Wright, 2005):

$$P_0 = \frac{\sum_{j=1}^J n_{jj}}{n_{..}} \quad [3]$$

$$P_c = \frac{\left(\frac{n_{.1} \times n_{1.}}{n_{..}}\right) + \dots + \left(\frac{n_{.j} \times n_{j.}}{n_{..}}\right)}{n_{..}} \quad [4]$$

$$\kappa = \frac{P_0 - P_c}{1 - P_c} \quad [5]$$

Dimana P_0 : observed proportion dan P_c : expected proportion.

4. Variable Importance Measure

Variable importance measure (VIM) dari OF untuk kovariat j adalah sebagai berikut (Hornung, 2020):

$$VI_j = \frac{1}{B_{ntree}} \sum_{b=1}^{B_{ntree}} [Err(y_{OOB,b,j}, \hat{y}_{OOB,b,j}) - Err(y_{OOB,b,j}, \hat{y}_{OOB,b,j})] \quad [6]$$

dengan:

$y_{OOB,b,j}$ adalah vektor nilai kelas dari data OOB pohon b dari f_{final} ,

$\hat{y}_{OOB,b,j}$ adalah prediksi dari nilai kelas dari data OOB pohon b dari f_{final} yang didapat setelah melakukan permutasi acak terhadap nilai-nilai kovariat j dalam data OOB pohon b ,

$\hat{y}_{OOB,b,j}$ adalah prediksi dari nilai kelas dari data OOB pohon b dari f_{final} tanpa melakukan permutasi terhadap nilai-nilai kovariat j ,

$Err(\{a_1, \dots, a_M\}, \{b_1, \dots, b_M\}) = (\frac{1}{M}) \sum_{m=1}^M I(a_m \neq b_m)$, yaitu galat misklasifikasi yang digunakan sebagai fungsi galat dalam VIM permutasi.

Metode Analisis

Data yang digunakan dalam penelitian ini adalah data sekunder mengenai Indeks Desa Membangun di Kabupaten Tasikmalaya dan Kabupaten Indramayu tahun 2021 yang bersumber dari Publikasi Direktorat Jenderal Pembangunan Desa dan Perdesaan Kementerian Desa, Pembangunan Daerah Tertinggal dan Transmigrasi serta data Potensi Desa tahun 2021 yang bersumber dari Badan Pusat Statistik Provinsi Jawa Barat. Adapun banyaknya desa di kabupaten Tasikmalaya yang menjadi objek penelitian ada sebanyak 351 desa dan banyaknya desa di kabupaten Indramayu yang menjadi objek penelitian ada sebanyak 309 desa. Metode klasifikasi yang dilakukan adalah metode hutan ordinal dan Naïve hutan ordinal. Proses analisis menggunakan R dengan *package*

ordinalForest (Hornung,2019) dan *package caret* (Khun et al., 2020)

Adapun peubah yang digunakan didasarkan pada indikator kemandirian desa (Sriningsih et al., 2020) yaitu (a) Ketahanan Sosial dalam Kemandirian Desa memiliki dua indikator yaitu, fasilitas pendidikan, dan fasilitas kesehatan, (b) Ketahanan Ekonomi menggunakan jumlah koperasi dan Badan Usaha Milik Desa (BUMDES) yang dimiliki oleh desa dan (c) Ketahanan Ekologi (Lingkungan) dengan menggunakan jumlah penduduk yang terdampak bencana di tahun sebelumnya. Selain ketiga indikator tersebut ditambahkan pula indikator terkait profil dari perangkat desa (kepala desa dan sekretaris desa). Secara keseluruhan pada penelitian ini menggunakan 17 peubah bebas dan 1 peubah respons, seperti pada tabel berikut:

Tabel 2. Daftar Peubah yang Digunakan

Peubah	Nama Peubah
	Jumlah penduduk terdampak bencana (X1)
	Sumber penghasilan utama (X2)
	Jumlah pengguna listrik (X3)
	Jarak terdekat ke SMA (X4)
	Jarak terdekat ke rumah sakit (X5)
	Jarak terdekat ke Puskesmas (X6)
	Jarak terdekat ke Poskesdes (X7)
	Jumlah posyandu (X8)
Peubah Bebas	Jarak kantor desa ke kantor bupati (X9)
	Jumlah koperasi (X10)
	Jumlah Badan Usaha Milik Desa (BUMDES) (X11)
	Jumlah Peraturan Desa di tahun 2020 (X12)
	Jumlah keluarga penerima BLT (3 bulan Pertama) (X13)
	Umur kepala desa (X14)
	Pendidikan terakhir kepala desa (X15)
	Umur sekretaris desa (X16)
	Pendidikan terakhir sekretaris desa (X17)

Peubah	Nama Peubah
Peubah Respons	Status IDM (Y)

Ketiga indikator kemandirian desa dijabarkan ke dalam 12 peubah bebas, dan profil desa dijabarkan ke dalam lima peubah bebas dengan pembagian sebagai berikut:

1. Ketahanan Sosial: X3, X4, X5, X6, X7, dan X8
2. Ketahanan Ekonomi : X2, X9, X10, X11, dan X13
3. Ketahanan Ekologi (Lingkungan) : X1
4. Profil Desa: X12, X14, X15, X16, dan X17

Adapun tahapan analisis data sebagai berikut:

1. Sinkronisasi data IDM 2021 dan data potensi desa 2021 untuk kabupaten Tasikmalaya dan kabupaten Indramayu
2. Eksplorasi data
3. Membagi data latih dan data uji dengan proporsi 75%:25%
4. Melakukan pemodelan menggunakan metode Naïve hutan ordinal dan hutan ordinal dengan banyaknya $nsets=1000$, $ntreeperdiv=100$, dan $ntreefinal=5000$
5. Evaluasi model dengan melihat tingkat akurasi dan nilai kappa pada setiap model
6. Menentukan peubah-peubah penting / VIM dari setiap model yang terbentuk
7. Tahapan 3-6 diulang sebanyak 100 ulangan guna mendapatkan hasil yang konsisten

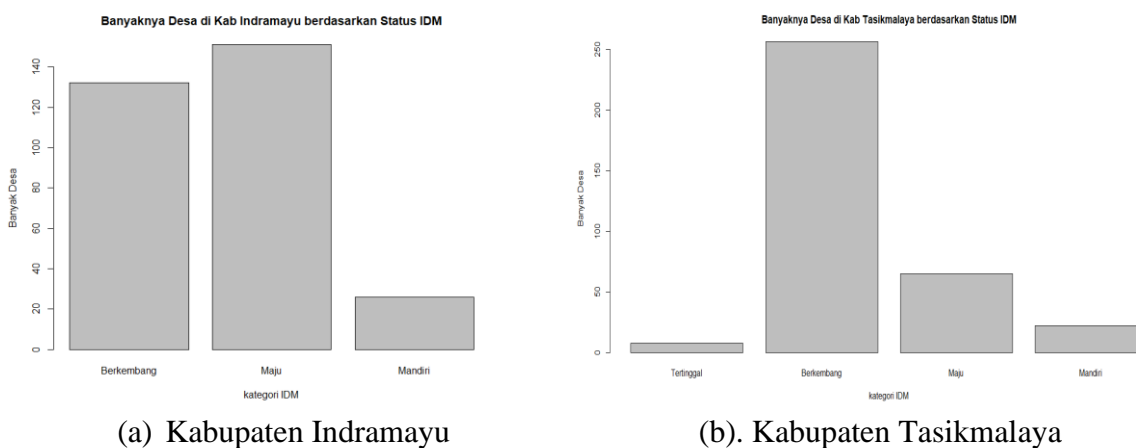
8. Menganalisis rata-rata tingkat akurasi dan nilai kappa dari hasil 100 ulangan baik secara deskriptif maupun inferensi.
9. Menganalisis peubah-peubah penting dari model terbaik yang dipilih di setiap daerah
10. Mengkaji peubah penjelas yang memiliki tingkat kepentingan tertinggi terhadap peubah respons dari model yang sudah terpilih.

HASIL DAN PEMBAHASAN

Tahapan awal dalam penelitian ini adalah melakukan eksplorasi data terhadap peubah respons dan peubah bebas. Hasilnya sebagai berikut:

Berdasarkan pada Gambar 1 dapat dilihat bahwa pada tahun 2021 status desa di Kabupaten Tasikmalaya didominasi berstatus IDM berkembang dan masih terdapat desa dengan status IDM tertinggal. Hal tersebut mengindikasikan perlunya percepatan pembangunan desa guna diharapkan desa-desa di Kab Tasikmalaya mempunyai status IDM maju dan berkembang. Sedangkan status desa di Kab Indramayu didominasi berstatus IDM maju dan berkembang, hal ini menjadi kontradiktif dengan rata-rata persentase penduduk miskin di Kabupaten Indramayu yang masih tinggi, yaitu sebesar 13.82%.

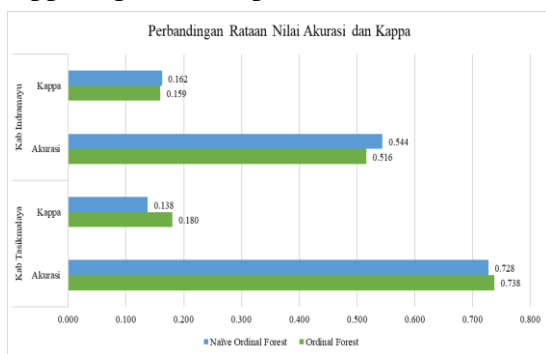
Selanjutnya proses klasifikasi status IDM akan didasarkan pada metode OF dan metode Naïve OF. Perbedaan kedua metode ini adalah terdapat pada proses nilai skor yang digunakan. Pada metode Naïve OF nilai skor akan di dasarkan pada pelabelan



Gambar 1. . Banyaknya Desa berdasarkan Status IDM

setiap kategorinya (Hornung, 2020), sehingga pada penelitian ini nilai skor pada metode Naïve OF akan disesuaikan dengan pelabelan pada status IDM yaitu 1: Sangat Tertinggal, 2: Tertinggal, 3: Berkembang, 4: Maju, dan 5: Mandiri. Adapun *hyperparameter* yang digunakan pada kedua metode tersebut adalah $nsets=1000$, $ntreeperdiv=100$, dan $ntreefinal=5000$. Proses analisis kedua metode diulang sebanyak 100 kali ulangan untuk melihat konsistensi *output* yang dihasilkan. Selanjutnya dilakukan eksplorasi baik secara deskriptif maupun inferensi terhadap nilai akurasi dan nilai kappa yang dihasilkan untuk setiap ulangan.

Parameter yang optimal dan model yang baik didapat ketika menghasilkan rata-rata tingkat akurasi dan nilai kappa yang tinggi. Rata-rata nilai akurasi dan nilai kappa dapat dilihat pada Gambar 2.



Gambar 2. Perbandingan Rata-Rata Nilai Akurasi dan Nilai Kappa

Berdasarkan pada Gambar 2 dapat dilihat bahwa di Kabupaten Tasikmalaya metode OF menghasilkan rata-rata akurasi sebesar 73.8% , rata-rata akurasi pada metode Naïve OF sebesar 72.8% dan rata-rata nilai kappa pada metode OF sebesar 0.18 sedangkan metode Naïve OF sebesar 0.138. Secara deskriptif dapat terlihat bahwa untuk kabupaten tasikmalaya metode OF memiliki tingkat akurasi dan nilai kappa yang lebih tinggi dibandingkan dengan metode Naïve OF dalam melakukan klasifikasi status IDM berdasarkan 17 peubah bebas yang digunakan.

Di daerah kabupaten Indramayu rata-rata akurasi pada metode OF sebesar 51.6 % , rata-rata akurasi pada metode Naïve OF sebesar 54.4% dan rata-rata nilai kappa pada metode OF sebesar 0.159 sedangkan

metode Naïve OF sebesar 0.162. Sehingga secara deskriptif didapatkan model Naïve OF memiliki tingkat akurasi dan nilai kappa yang lebih tinggi dalam melakukan klasifikasi status IDM berdasarkan 17 peubah bebas yang digunakan

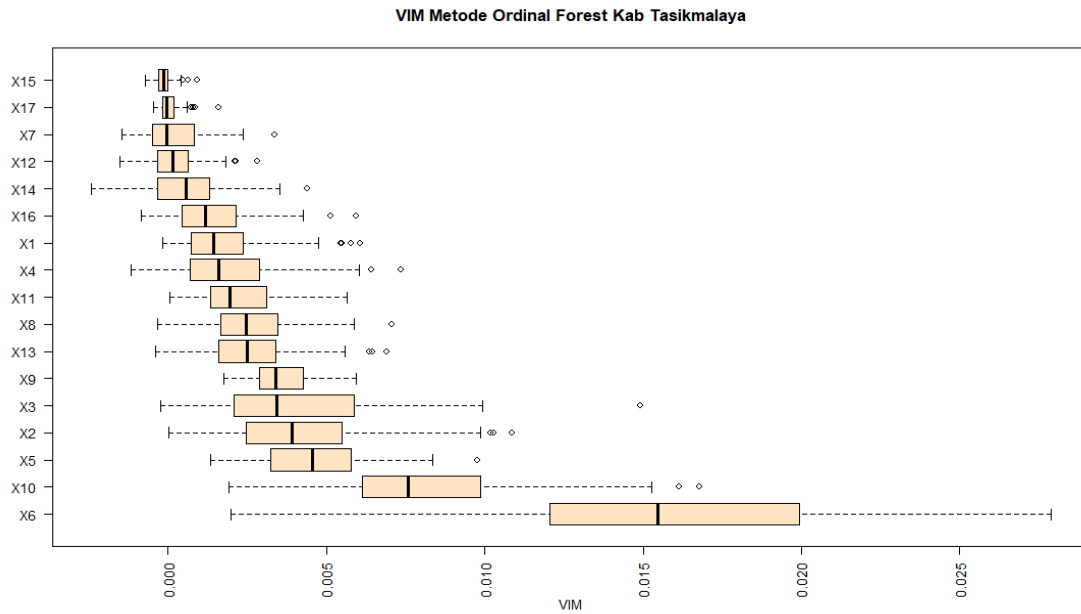
Untuk dapat memperjelas metode mana yang lebih baik, maka dilakukan pengujian hipotesis terkait apakah terdapat perbedaan rata-rata akurasi dan rata-rata nilai kappa di antara metode OF dan Naïve OF di kedua daerah ini. Hasil pengujian hipotesis ditampilkan pada Tabel 3 berikut:

Tabel 3. Hasil Pengujian Hipotesis Perbedaan Akurasi dan Kappa diantara Metode OF dan Naïve OF

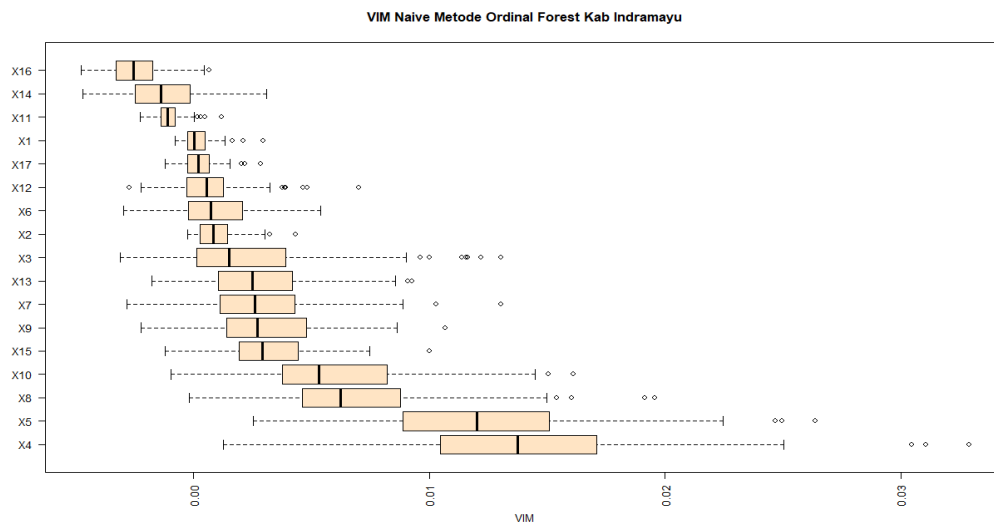
		<i>T-test</i>	<i>P-value</i>
Kab Tasikmalaya	Akurasi	5.970	3.72E-08
	Kappa	7.8251	5.68E-12
Kab Indramayu	Akurasi	-5.415	4.30E-07
	Kappa	-0.3912	0.6965

Hasil pengujian hipotesis terkait metode mana yang lebih baik dalam mengklasifikasikan status IDM berdasarkan 17 peubah yang dianalisis di Kabupaten Tasikmalaya dan Kabupaten Indramayu terlihat pada tabel 3. Di kabupaten Tasikmalaya untuk ukuran akurasi dan nilai kappa memiliki nilai $p\text{-value} < \alpha$ (0.05) yang mana memperlihatkan bahwa tolak H_0 yakni terdapat perbedaan nilai rata-rata akurasi dan rata-rata nilai kappa untuk metode OF dan Naïve OF, artinya bahwa kedua metode klasifikasi tersebut menghasilkan prediksi status IDM yang berbeda. Hal ini sejalan dengan eksplorasi sebelumnya bahwa metode OF memiliki tingkat akurasi dan nilai kappa yang lebih tinggi.

Hal berbeda terjadi di daerah kabupaten Indramayu, pada nilai akurasi memiliki nilai $p\text{-value} < \alpha$ (0.05) yang mana memperlihatkan bahwa tolak H_0 yakni terdapat perbedaan nilai rata-rata akurasi untuk metode OF dan Naïve OF, sedangkan untuk nilai kappa nilai $p\text{-value} > \alpha$ (0.05) yang mana memperlihatkan bahwa gagal tolak H_0 yakni tidak ada perbedaan rata-rata nilai kappa untuk metode OF dan



Gambar 3. VIM dengan metode Hutan ordinal pada klasifikasi status IDM di Kabupaten Tasikmalaya



Gambar 4. VIM dengan metode Naïve Hutan ordinal pada klasifikasi status IDM di Kabupaten Indramayu

Naïve OF. Hanya saja pada proses eksplorasi sebelumnya terlihat bahwa metode Naïve OF memiliki tingkat akurasi dan nilai kappa yang lebih tinggi.

Pada metode yang memberikan kinerja lebih baik di kedua daerah tersebut, akan dicari peubah-peubah yang dirasakan penting dalam proses klasifikasi status IDM, untuk kabupaten tasikmalaya perhitungan VIM akan didasarkan pada metode OF sedangkan kabupaten Indramayu perhitungan VIM akan didasarkan pada metode Naïve OF. Dari

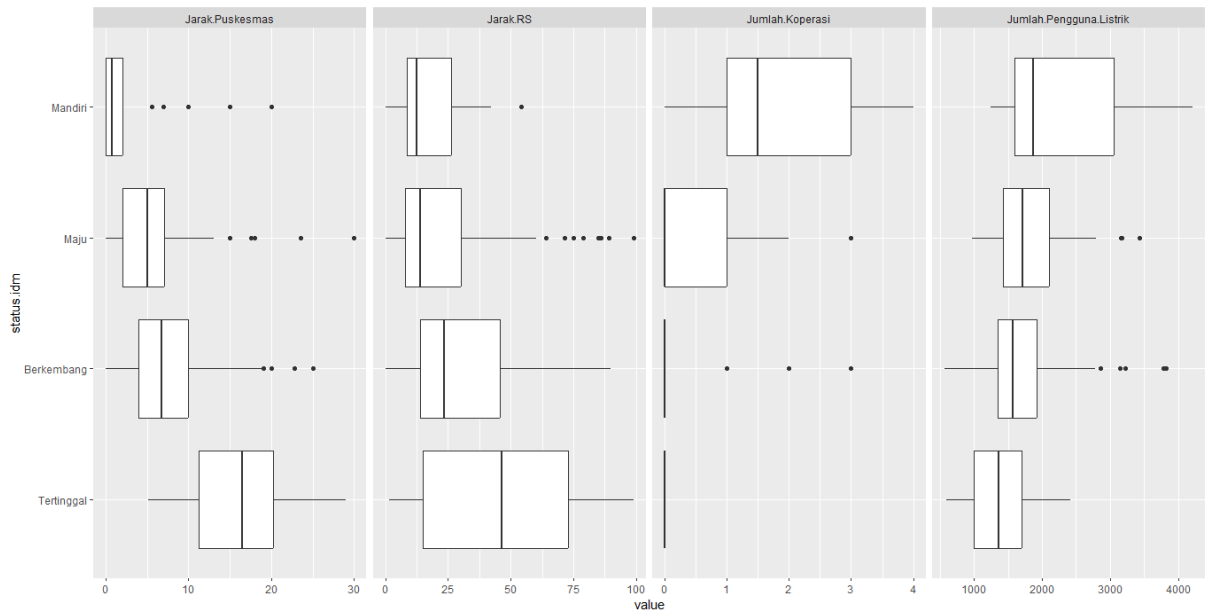
100 ulangan yang dilakukan, gambaran nilai VIM ditampilkan pada Gambar 3 dan Gambar 4.

Gambar 3 dan Gambar 4 menunjukkan Box plot dari hasil 100 ulangan nilai VIM dalam menentukan peubah penting dalam proses klasifikasi status IDM. Lima peubah teratas untuk kab tasikmalaya memiliki nilai VIM lebih besar dari 0.0040 yaitu (1) Jarak terdekat ke Puskesmas (X6), (2) Jumlah koperasi (X10), (3) Jarak terdekat ke rumah sakit

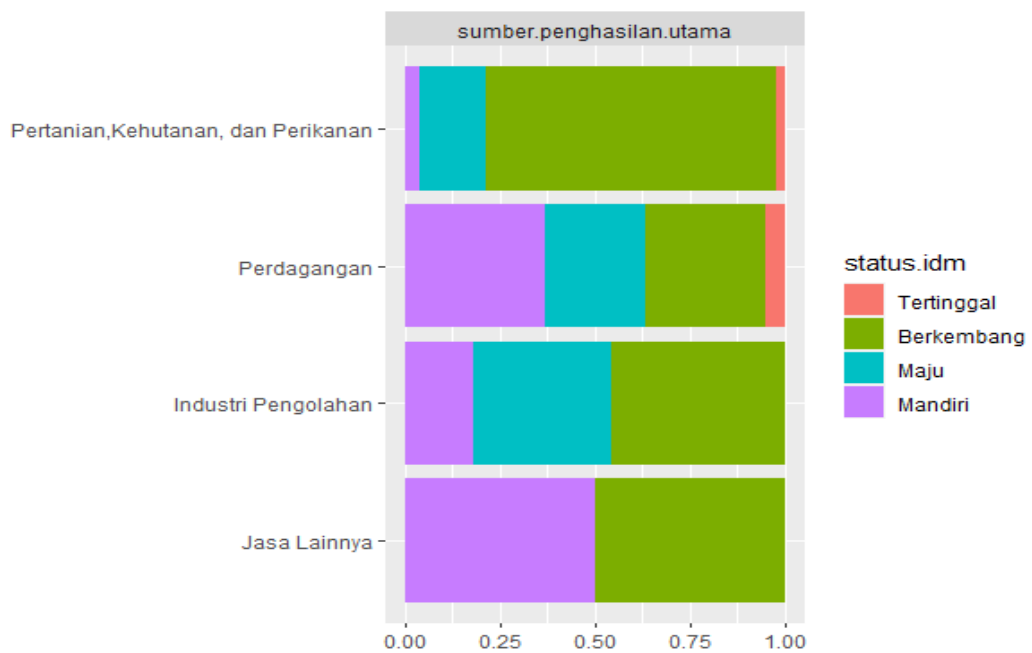
(X5), (4) Sumber penghasilan utama (X2), dan (5) Jumlah pengguna listrik (X3).

Sedangkan lima peubah teratas untuk kab Indramayu memiliki nilai VIM lebih besar dari 0.0032 yaitu (1) Jarak terdekat ke SMA (X4), (2) Jarak terdekat ke rumah sakit (X5), (3) Jumlah posyandu (X8), (4) Jumlah koperasi (X10), dan (5) Pendidikan terakhir kepala desa (X15). Proses

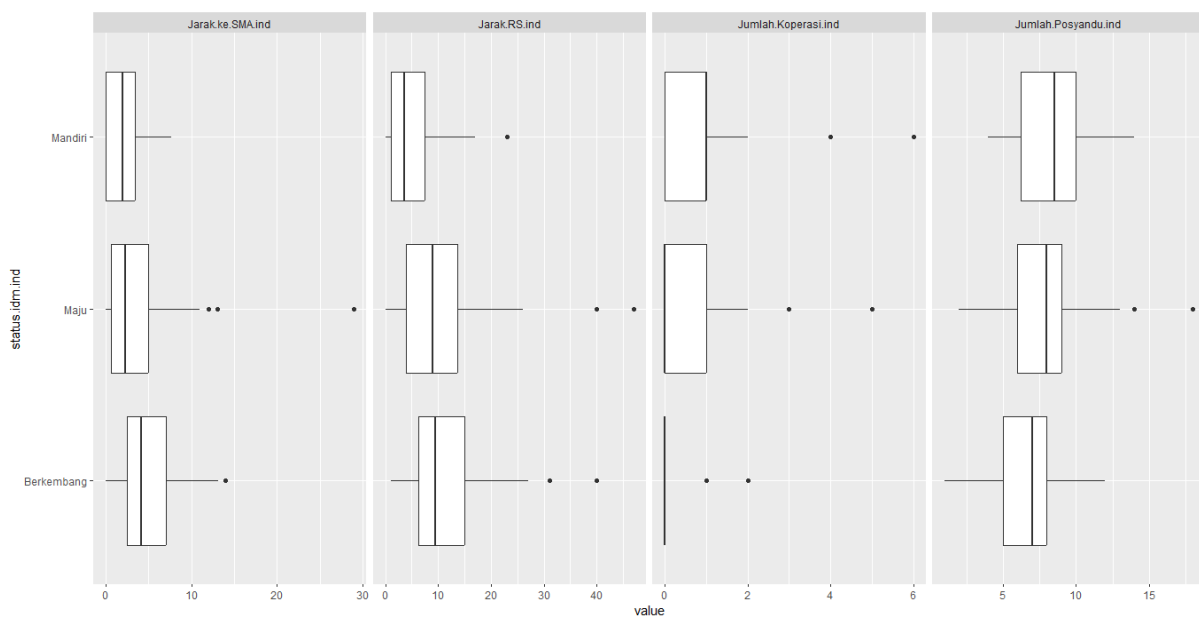
eksplorasi terhadap kelima peubah tersebut jika dikaitkan dengan klasifikasi status IDM sendiri dapat dijelaskan pada Gambar 5-7 berikut:



Gambar 5. Box plot Peubah yang memiliki nilai VIM tertinggi di Kabupaten Tasikmalaya



Gambar 6. Bar plot Peubah yang memiliki nilai VIM tertinggi di Kabupaten Tasikmalaya



Gambar 7. *Box plot* Peubah yang memiliki nilai VIM tertinggi di Kabupaten Indramayu

Lima peubah dengan nilai VIM tertinggi dalam proses klasifikasi status IDM di daerah kabupaten tasikmalaya dapat dideskripsikan seperti pada gambar 5 dan gambar 6. Pada peubah Jarak terdekat ke Puskesmas dan Jarak terdekat ke rumah sakit mengindikasikan bahwa semakin jauh jarak desa ke puskesmas atau rumah sakit maka desa tersebut akan diklasifikasikan ke dalam kategori desa tertinggal dan sebaliknya desa dengan kategori mandiri memiliki jarak yang paling dekat di antara desa dengan puskesmas atau rumah sakit. Hal yang berbeda untuk peubah Jumlah koperasi dan Jumlah pengguna listrik memperlihatkan bahwa semakin sedikit jumlah koperasi dan jumlah pengguna listrik di suatu desa maka desa tersebut akan diklasifikasikan ke dalam kategori desa tertinggal dan sebaliknya semakin banyak jumlah koperasi dan jumlah pengguna listrik di suatu desa maka desa tersebut akan diklasifikasikan ke dalam kategori desa mandiri. Sedangkan peubah Sumber penghasilan utama di kabupaten tasikmalaya didominasi oleh penduduk yang sumber penghasilan utamanya ada di bidang pertanian, kehutanan, dan perikanan serta masuk pada klasifikasi desa berkembang.

Lima peubah dengan nilai VIM tertinggi dalam proses klasifikasi status IDM di daerah kabupaten Indramayu dapat

dideskripsikan seperti pada gambar 7. Pada peubah Jarak terdekat ke SMA dan Jarak terdekat ke rumah sakit mengindikasikan bahwa semakin jauh jarak desa ke SMA atau rumah sakit maka desa tersebut akan diklasifikasikan ke dalam kategori desa berkembang dan sebaliknya desa dengan kategori mandiri memiliki jarak yang paling dekat di antara desa dengan SMA atau rumah sakit. Hal yang berbeda untuk peubah Jumlah koperasi dan Jumlah posyandu memperlihatkan bahwa semakin sedikit jumlah koperasi dan Jumlah posyandu di suatu desa maka desa tersebut akan diklasifikasikan ke dalam kategori desa berkembang dan sebaliknya semakin banyak Jumlah koperasi dan Jumlah posyandu di suatu desa maka desa tersebut akan diklasifikasikan ke dalam kategori desa mandiri.

Dari sisi peubah Pendidikan terakhir kepala desa memperlihatkan bahwa di kabupaten Indramayu sebesar 37.22 % kepala desa memiliki Pendidikan terakhir SMU/ sederajat dan 33.66% kepala desa memiliki Pendidikan terakhir Diploma IV/S1. Jika dilakukan pengujian terhadap asosiasi antara tingkat Pendidikan terakhir kepala desa terhadap status IDM menunjukkan bahwa nilai Tau Kendall sebesar 0.1103397 dan nilai p -value (0.03053) $< \alpha$ (0.05). hal ini

mengindikasikan bahwa memang terdapat asosiasi diantara kedua peubah tersebut.

Jika dibandingkan kelima peubah dengan nilai VIM tertinggi baik untuk daerah Kabupaten Tasikmalaya maupun Kabupaten Indramayu ditampilkan pada Tabel 4 berikut:

Berdasarkan tabel 4 terlihat kesamaan peubah yang dipentingkan dalam proses klasifikasi status IDM di daerah Kabupaten Tasikmalaya dan Kabupaten Indramayu adalah peubah Jumlah Koperasi dan Jarak terdekat ke rumah sakit. Peran koperasi juga sangat berpengaruh terhadap kesejahteraan masyarakat dimana menurut Undang-Undang No 25 tahun 1992 dijelaskan bahwa peran koperasi diantaranya (a) membangun dan mengembangkan potensi dan kemampuan ekonomi anggotanya dan masyarakat, (b) berperan aktif dalam meningkatkan kualitas manusia dan masyarakat, (c) memperkuat perekonomian rakyat, dan (d) mewujudkan dan mengembangkan perekonomian nasional (Zulhartati, 2010). Hal ini memperlihatkan bahwa kemajuan suatu desa sangat ditentukan dari aspek ekonomi yaitu jumlah koperasi yang dimiliki di desa dan juga dari sisi sarana dan pra sarana di bidang Kesehatan yakni jarak dari desa ke rumah sakit terdekat.

KESIMPULAN DAN SARAN

Kesimpulan

Metode OF dan Naïve OF dapat digunakan sebagai metode klasifikasi berbasis pohon yang diaplikasikan pada peubah respons ordinal. Dalam konteks klasifikasi status IDM berdasarkan faktor sosial, ekonomi, ekologi dan juga profil dari perangkat desa, di daerah kabupaten Tasikmalaya metode OF memberikan kinerja yang lebih baik dibandingkan metode Naïve OF akan tetapi untuk daerah kabupaten Indramayu metode Naïve OF tampak lebih baik dibandingkan metode OF. Selain itu dari sisi peubah yang dipentingkan dalam proses klasifikasi status IDM di daerah Kabupaten Tasikmalaya dan Kabupaten Indramayu adalah peubah

Jumlah Koperasi dan Jarak terdekat ke rumah sakit.

Saran

Berdasarkan analisis yang telah dilakukan maka saran yang dapat penulis berikan adalah :

1. Dalam penelitian yang telah dilakukan terdapat perbedaan kinerja di antara metode OF dan Naïve OF hal ini didasarkan pada dasar pelabelan yang dilakukan pada peubah respons, selanjutnya dapat dikembangkan model lain yang berbasis pada pemodelan berbasis peubah respons ordinal.
2. Bagi peneliti selanjutnya dapat menambahkan peubah dalam indikator ketahanan ekologi, karena dalam penelitian ini hanya terbatas menggunakan peubah jumlah penduduk yang terdampak bencana di tahun sebelumnya saja dalam indikator tersebut sehingga harapannya dapat meningkatkan nilai akurasi model yang dihasilkan.

Dalam rangka mewujudkan kemandirian desa-desa di Indonesia maka pemerintah setempat dapat mengoptimalkan fasilitas kesehatan desa dan juga keberadaan koperasi desa.

DAFTAR PUSTAKA

- Agustine, P. Y. (2019). Penerapan Random Forest Ordinal dalam Menentukan Mutu Penciri Akreditasi SMA berdasarkan Indikator Pemetaan Mutu Pendidikan.
- Badan Pusat Statistik Provinsi Jawa Barat. (n.d.). Retrieved March 18, 2023, from <https://jabar.bps.go.id/publication/2022/08/29/132e723cf83a810777bc74a1/perkembangan-tingkat-kemiskinan-provinsi-jawa-barat-maret-2022.html>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324/METRICS>
- Dewi, N. K., Syafitri, U. D., & Mulyadi, S. Y. (2011). Penerapan Metode Random Forest dalam Driver

- Analysis. *FORUM STATISTIKA DAN KOMPUTASI*, 16(1), 35–43. <https://jurnal.ipb.ac.id/index.php/statistika/article/view/5443>
- Fitria, M. C. N., Debatara, N. N., & Rizki, S. W. (2022). Classification of Village Status in Landak Regency Using C5.0 Algorithm. *Tensor: Pure and Applied Mathematics Journal*, 3(1), 33–42. <https://doi.org/10.30598/TENSORVOL3ISS1PP33-42>
- Hanibal, H. (2015). Indeks Desa Membangun.
- Hornung, R. (2017). Ordinal Forests.
- Hornung, R. (2020). Ordinal Forests. *Journal of Classification*, 37(1), 4–17. <https://doi.org/10.1007/S00357-018-9302-X/METRICS>
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. [Http://Dx.Doi.Org/10.1198/106186006X133933](http://Dx.Doi.Org/10.1198/106186006X133933), 15(3), 651–674. <https://doi.org/10.1198/106186006X133933>
- Irawan, H. (2019). Faktor-Faktor Rumah Tangga yang Mencirikan Tingkat Kerawanan Pangan.
- Janitza, S., Tutz, G., & Boulesteix, A. L. (2016). Random forest for ordinal responses: Prediction and variable selection. *Computational Statistics & Data Analysis*, 96, 57–73. <https://doi.org/10.1016/J.CSDA.2015.10.005>
- Narulita, R., & Kudus, A. (2022). Penerapan Gaussian Mixture Model pada Klasterisasi Desa di Kalimantan Barat berdasarkan Indeks Desa Membangun Tahun 2021. *Bandung Conference Series: Statistics*, 2(2), 226–233. <https://doi.org/10.29313/BCSS.V2I2.3877>
- Nisa, Z. I. (2019). Identifikasi Faktor-Faktor yang Memengaruhi Prestasi Mahasiswa Menggunakan Regresi Logistik Ordinal dan Random Forest Ordinal (Studi Kasus Mahasiswa FMIPA IPB Angkatan 2015-2017).
- Setyowati, E. (2019). Tata Kelola Pemerintahan Desa Pada Perbedaan Indeks Desa Membangun (Idm): Studi Tiga Desa Di Kabupaten Malang. In *JISPO* (Vol. 9, Issue 2).
- Sim, J., & Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, 85(3), 257–268. <https://doi.org/10.1093/PTJ/85.3.257>
- Sriningsih, S., Astuti, E., Ismiwati, B., Ekonomi, F., & Bisnis, D. (2020). Implementasi PERMENDESAPDTRANS NO. 2 Tahun 2016 Terkait Status Desa di Desa Sukarara Lombok Tengah. In *Jurnal Kompetitif: Media Informasi Ekonomi Pembangunan* (Vol. 6, Issue 1). <http://www.keuangandes.com>.
- Zulhartati, S. (2010). Peranan Koperasi Dalam Perekonomian Indonesia. *Guru Membangun*, 25(3). <https://doi.org/10.26418/GM.V25I3.233>

PEMODELAN *MIXED GEOGRAPHICALLY WEIGHTED REGRESSION-SPATIAL AUTOREGRESSIVE* (MGWR-SAR) PADA KASUS HIV DI INDONESIA

Anik Djuraidah¹, Rahma Anisa¹, Arna Ristiyanti Tarida¹, Muftih Alwi Aliu¹, Cintia Septemberini¹, Yufan Putri Astrini¹, Gusti Tasya Meilania¹

¹Departement of Statistics, IPB University, Bogor, Indonesia
e-mail: anikdjuraidah@apps.ipb.ac.id

Received: July 1, 2023; Revised: November 16, 2023; Accepted: December 29, 2023; Published: December 31, 2023

Abstrak

Pada umumnya regresi spasial memodelkan salah satu efek spasial, yaitu dependensi atau heterogenitas spasial. Untuk efek dependensi spasial, model yang telah sering dipakai mengikuti taksonomi Elhost, dengan dependensi spasial terdapat pada peubah respon, prediktor, atau galat. Sedangkan untuk efek heterogenitas spasial umumnya menggunakan model regresi terboboti geografis (GWR) atau bila terdapat prediktor yang global digunakan regresi terboboti geografis campuran (MGWR). Data yang digunakan pada penelitian ini adalah kasus Human Immunodeficiency Virus (HIV) per 100.000 penduduk sebagai peubah respon, dan populasi kunci, kasus positif pada ibu hamil, pasien tuberkulosis, tingkat kemiskinan, tingkat pengangguran sebagai prediktor. Pada data yang digunakan terdapat efek dependensi dan heterogenitas spasial. MGWR-SAR merupakan model yang dapat digunakan bila pada data terdapat kedua efek spasial tersebut. Penelitian ini bertujuan untuk menentukan faktor-faktor yang mempengaruhi kasus HIV di kabupaten/kota di Indonesia menggunakan model spasial. Hasil penelitian menunjukkan bahwa model MGWR-SAR $(0, k_c, k_v)$ merupakan model terbaik yang terpilih berdasarkan nilai RMSE terkecil yaitu sebesar 0,54066. Prediktor populasi kunci memiliki pengaruh global dan signifikan. Prediktor lain memiliki pengaruh secara lokal adalah kasus positif pada ibu hamil, pasien tuberkulosis, tingkat kemiskinan, dan tingkat pengangguran.

Kata kunci: GWR, HIV, MGWR, MGWR-SAR, SAR

Abstract

In general, spatial regression is used to model one of the spatial effects, namely spatial dependency or heterogeneity. For the effects of spatial dependencies, the models that have been used frequently follow Elhost's taxonomy, with the spatial dependencies being on the response, predictor, or error. Whereas for the effect of spatial heterogeneity generally use geographically weighted regression models (GWR) or if there are global predictors use mixed geographically weighted regression (MGWR). The data used in this study are cases of Human Immunodeficiency Virus (HIV) per 100,000 population as a response variable, and key populations, positive cases in pregnant women, tuberculosis patients, poverty rate, and unemployment rate as predictors. In the data used, there are spatial dependencies and heterogeneity. The MGWR-SAR is a model that can be used if the data has both spatial effects. This study aims to determine the factors influencing HIV cases in districts/cities in Indonesia using a spatial model. The results showed that the combined model of GWR and spatial autoregressive regression (SAR) was the best model. The research results indicate that the MGWR-SAR model $(0, k_c, k_v)$ is the best model selected based on the smallest RMSE value, which is 0.54066. Key population explanatory variables have a global and significant influence. Other explanatory variables that have local influence are positive cases in pregnant women, tuberculosis patients, poverty rates, and unemployment rates.

Keywords: GWR, HIV, MGWR, MGWR-SAR, SAR

PENDAHULUAN

Data spasial adalah data yang berorientasi geografis, mengacu pada sistem koordinat tertentu, dan memiliki dua bagian penting yang membedakannya dengan data lain, yaitu informasi lokasi dan informasi deskriptif. Anselin (1988) menyatakan bahwa informasi lokasi menghasilkan dua efek spasial, yaitu dependensi spasial dan heterogenitas spasial. Jika data yang mengandung efek spasial dianalisis menggunakan metode kuadrat terkecil (MKT) maka model yang dihasilkan akan menjadi kurang tepat karena asumsi kebebasan dan kehomogenan galat yang tidak terpenuhi. Oleh karena itu, diperlukan metode analisis statistika yang mempertimbangkan efek spasial (Comber dan Brunson, 2021), salah satunya yaitu model *Spatial Autoregressive* (SAR).

SAR yang juga dikenal sebagai *spatial lag* adalah analisis regresi spasial dengan efek dependensi spasial pada peubah respon. Menurut Anselin (1988), model SAR merupakan model regresi spasial yang dapat memperhitungkan efek dependensi spasial. Namun, Shekhar *et al.* (2018) menyatakan bahwa model SAR tidak cocok digunakan pada kasus heterogenitas spasial, sehingga perlu dilakukan pemodelan dengan regresi terboboti geografis (*Geographically Weighted Regression/ GWR*).

Pendekatan GWR merupakan modifikasi dari model regresi klasik yang dapat memperhitungkan efek heterogenitas spasial. Model GWR menghasilkan penduga parameter lokal untuk setiap lokasi pengamatan dengan setiap parameter dihitung pada setiap titik lokasi geografis (Lu *et al.*, 2014). Dalam analisisnya, GWR menggunakan matriks pembobot yang tergantung pada kedekatan lokasi dengan fungsi kernel.

Faktanya, tidak semua parameter regresi dalam model GWR bervariasi secara spasial, beberapa parameter mungkin tidak signifikan secara spasial. Model GWR kemudian dikembangkan menjadi model

Mixed Geographically Weighted Regression (MGWR) yang menggabungkan model regresi linier dan GWR. Dalam MGWR, beberapa parameter GWR tetap konstan di semua lokasi, sementara yang lain bervariasi pada lokasi pengamatan. Hal ini menghasilkan penduga parameter global dan lokal dalam model MGWR (Fotheringham, Brunson, dan Charlton., 2002).

Pada suatu data kadang dijumpai terdapat efek dependensi dan keragaman spasial. (Geniaux dan Martinetti, 2018) mengembangkan MGWR dengan penambahan model dependensi SAR yang dikenal dengan *Mixed Geographically Weighted Regression-Spatial Autoregressive* (MGWR-SAR). Beberapa model yang dikembangkan diberi notasi MGWR-SAR $(0, k_c, k_v)$, dan MGWR-SAR $(1, k_c, k_v)$. MGWR-SAR $(0, k_c, k_v)$ menunjukkan bahwa model mempunyai koefisien dependensi (ρ) yang sama di semua lokasi, sedangkan MGWR-SAR $(1, k_c, k_v)$ menunjukkan koefisien dependensi (ρ) yang berbeda-beda antar lokasi. Konstanta k_c menunjukkan banyaknya parameter regresi global dan k_v menunjukkan banyaknya parameter regresi lokal.

Tujuan dari penelitian ini yaitu menentukan model regresi spasial terbaik pada kasus *Human Immunodeficiency Virus* (HIV) tahun 2018. HIV merupakan virus yang menginfeksi sel-sel penting dalam sistem imun tubuh. Virus ini dapat ditularkan melalui kontak langsung dengan cairan tubuh yang terinfeksi, mulai dari transfusi darah sampai kontak seksual. Indonesia adalah negara dengan jumlah kasus HIV tertinggi ketiga di Asia Pasifik, hingga tahun 2018 jumlah penderita sebanyak 327.282 jiwa (Kemenkes RI, 2019). Sehubungan dengan tingginya angka tersebut, maka sangat penting bagi masyarakat Indonesia untuk meningkatkan kewaspadaan terhadap penyebaran dan penularan virus HIV. Oleh karena itu, perlu dilakukan pemodelan untuk mengetahui

faktor-faktor yang diduga berpengaruh terhadap penyebaran dan penularan HIV.

Penelitian terkait HIV sebelumnya dilakukan oleh El-Kautsar, Djuraidah dan Angraini (2022) menggunakan MGWR dengan fungsi pembobot *kernel Bisquare* untuk mengidentifikasi faktor pengaruh kasus HIV di Indonesia. Metode serupa pernah digunakan oleh Kusnandar, Debataraja dan Fitriani (2021) untuk analisis kualitas air di Kota Pontianak. Penelitian lain yang dilakukan oleh (Olakunde *et al.*, 2021) menggunakan model SAR untuk mengidentifikasi faktor-faktor risiko yang terkait dengan HIV dalam hal demografi, sosial, dan perilaku dengan mempertimbangkan efek spasial. Sementara itu, penelitian menggunakan model MGWR-SAR sebelumnya dilakukan oleh (Furková, 2021).

METODOLOGI

Tinjauan Referensi

1. GWR

GWR merupakan pengembangan dari model regresi global yang digunakan untuk memodelkan dan menganalisis parameter yang memiliki variabilitas spasial, sehingga nilai parameter regresi akan berbeda pada setiap lokasi (Fotheringham *et al.*, 2002).

Model GWR dapat ditulis sebagai berikut:

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

untuk $i = 1, 2, \dots, n; k = 1, 2, \dots, p; (u_i, v_i)$ adalah titik koordinat ke- i (bujur, lintang); y_i adalah peubah respon pada lokasi ke- i ; $\beta_k(u_i, v_i)$ adalah koefisien regresi pada lokasi ke- i ; x_{ik} adalah peubah prediktor ke- k pada lokasi ke- i ; dan ε_i merupakan galat untuk lokasi ke- i .

Fungsi pembobot pada GWR digunakan untuk memberikan bobot pada setiap lokasi menggunakan fungsi kernel. Lokasi yang lebih dekat memiliki bobot yang besar dalam pendugaan parameter daripada lokasi yang jauh (Anselin, 1988). Beberapa fungsi pembobot *kernel*, diantaranya yaitu *Kernel Gaussian*, *Exponential*, *Bisquare*, *Tricube*, dan *Adaptive* (Djuraidah, 2020). Dalam penelitian ini menggunakan fungsi *kernel*:

- Fungsi *Kernel Fixed Gaussian*

$$w_j(u_i, v_i) = \exp\left\{-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right\}$$

Dengan d_{ij} adalah jarak *Euclidean* antara lokasi (u_i, v_i) ke lokasi (u_j, v_j) , dimana:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$$

dan h adalah lebar jendela (*bandwidth*).

- Fungsi *Kernel Fixed Bisquare*

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right)^2, & \text{untuk } d_{ij} \leq h \\ 0, & \text{untuk } d_{ij} > h \end{cases}$$

- Fungsi *Kernel Adaptive Gaussian*

$$w_j(u_i, v_i) = \exp\left\{-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right\}$$

- Fungsi *Kernel Adaptive Bisquare*

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h_i}\right)^2\right)^2, & \text{untuk } d_{ij} \leq h_i \\ 0, & \text{untuk } d_{ij} > h_i \end{cases}$$

Bandwidth (h) dihitung melalui *Cross Validation* (CV), dengan rumus:

$$CV = \sum_{i=1}^n (y_i - \hat{y}_{(i)}(h))^2$$

Dengan $\hat{y}_{(i)}(h)$ merupakan *fitted value* dari y_i yang menghilangkan pengamatan di lokasi (u_i, v_i) dari proses *fitting*. Nilai h yang paling optimal ditentukan berdasarkan CV terkecil.

2. MGWR

MGWR-SAR memiliki notasi (i_ρ, k_c, k_v) dimana i_ρ dapat dinyatakan bahwa lag konstan ($\rho=0$) atau bervariasi secara spasial ($\rho=1$), k_c merupakan jumlah parameter regresi global, dan k_v adalah jumlah parameter regresi lokal. Model MGWR-SAR $(0, k_c, k_v)$ dan MGWR-SAR $(1, k_c, k_v)$ dituliskan sebagai berikut:

MGWR-SAR(0, k_c, k_v)

$$y_i = \rho W y_i + \beta_0(u_i, v_i) + \sum_{k=1}^q \beta_k(u_i, v_i)x_{ik} + \sum_{k=q+1}^p \beta_k x_{ik} + \varepsilon_i$$

MGWR-SAR(1, k_c, k_v)

$$y_i = \rho(u_i, v_i)W y_i + \beta_0(u_i, v_i) + \sum_{k=1}^q \beta_k(u_i, v_i)x_{ik} + \sum_{k=q+1}^p \beta_k x_{ik} + \varepsilon_i$$

$W y_i$ adalah matriks bobot spasial yang menghubungkan nilai peubah respon y_i di lokasi (u_i, v_i)

3. Evaluasi Signifikansi Parameter

Penelitian ini menggunakan uji *Wald* untuk menguji signifikansi parameter pemodelan spasial. Parameter β diuji dengan hipotesis:

$$H_0: \beta_j = 0$$

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

Pendekatan statistik pengujian *Wald*:

$$Wald_{\beta} = \frac{\hat{\beta}_j^2}{var(\hat{\beta}_j)}$$

Hipotesis untuk menguji parameter λ adalah:

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

$$\text{Statistik uji: } Wald_{\lambda} = \frac{\hat{\rho}^2}{var(\rho)}$$

Keputusan yang diambil adalah menolak H_0 ketika nilai statistik uji *Wald* melebihi nilai kritis $\chi^2_{(\alpha, 1)}$.

4. Metode Analisis

Data yang digunakan merupakan data sekunder hasil pemantauan perkembangan situasi HIV tahun 2018 dan peubah-peubah yang diduga berperan dalam memengaruhi peubah respon. Data diperoleh dari Kementerian Kesehatan, Badan Perencanaan Pembangunan Nasional (Bappenas), dan Badan Pusat Statistik (BPS). Jumlah lokasi amatan dengan total kasus HIV tidak nol sebanyak 390 kabupaten/kota dari seluruh Indonesia.

Peubah yang dijadikan acuan dalam penelitian ini yaitu kasus HIV per 100.000 penduduk sebagai peubah respon. Populasi kunci per 100.000 penduduk, kasus positif pada ibu hamil per 100.000 penduduk, pasien tuberkulosis per 100.000 penduduk, tingkat kemiskinan, dan tingkat pengangguran sebagai peubah prediktor. Metode yang digunakan adalah MGWR,

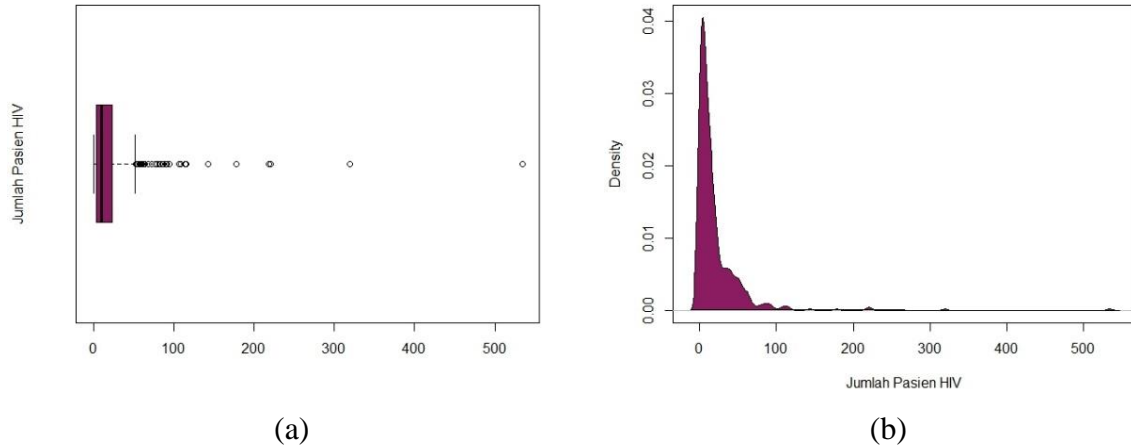
MGWR-SAR (0, k_c, k_v), dan MGWR-SAR (1, k_c, k_v). Berikut adalah urutan langkah-langkah yang dilakukan dalam penelitian ini.

1. Melakukan eksplorasi data untuk mengetahui pola sebaran peubah respon dan plot pencaran antara peubah respon dan prediktor.
2. Melakukan uji multikolinieritas dengan menggunakan *variance inflation factor* (VIF).
3. Menguji efek spasial dengan menerapkan uji dependensi spasial pada respon dan galat menggunakan uji *Robust Lagrange Multiplier* /RLM (Anselin, 1988) dan menguji efek keragaman spasial melalui uji *Breusch Pagan* /BP (Arbia, 2006).
4. Menetapkan matriks pembobot dengan membandingkan empat fungsi pembobot yang berbeda, yaitu *Fixed Gaussian*, *Fixed Bisquare*, *Adaptive Gaussian*, dan *Adaptive Bisquare*. Fungsi pembobot yang optimal ditentukan berdasarkan nilai *Akaike Information Criterion* (AIC) terkecil.
5. Melakukan pemodelan GWR dengan fungsi pembobot optimal untuk mengidentifikasi parameter yang berpengaruh secara global maupun lokal.
6. Menduga parameter dalam model MGWR dengan fungsi pembobot optimal.
7. Melakukan pendugaan parameter model MGWR-SAR (0, k_c, k_v) dan MGWR-SAR (1, k_c, k_v).
8. Membandingkan hasil model MGWR, MGWR-SAR (0, k_c, k_v), dan MGWR-SAR (1, k_c, k_v).
9. Menerapkan uji *Wald* untuk menguji signifikansi parameter.
10. Melakukan interpretasi pada hasil analisis.

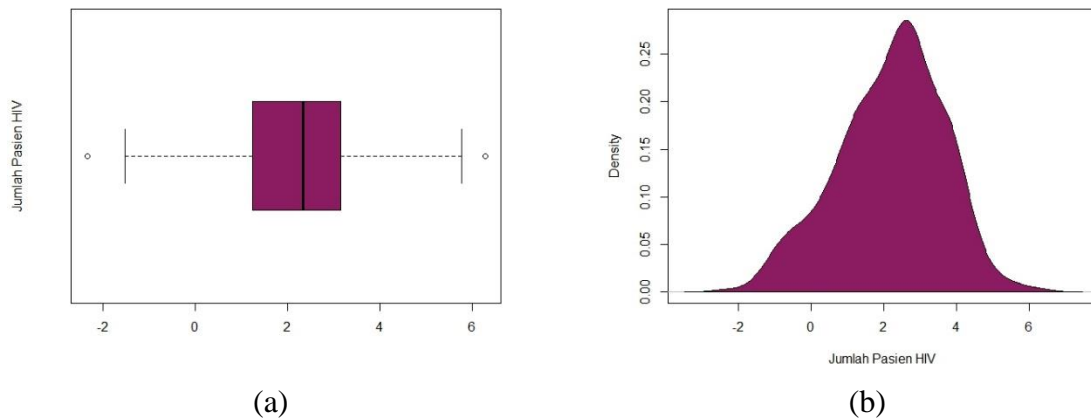
HASIL DAN PEMBAHASAN

Eksplorasi Data

Karakteristik umum data peubah respon sebelum transformasi ditunjukkan pada **Error! Reference source not found.** D ata menjulur ke kanan dan memiliki banyak



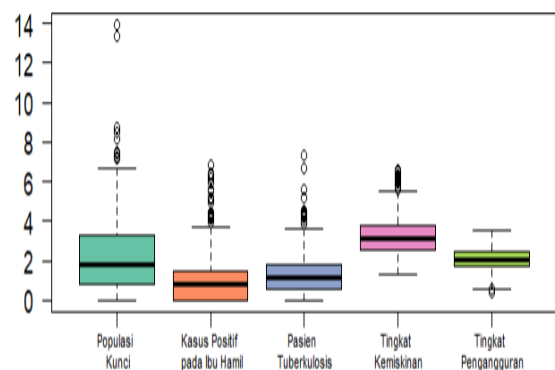
Gambar 1. Ekplorasi Data Kasus HIV (a) Diagram kotak garis, (b) Sebaran densitas



Gambar 2. Ekplorasi Data Hasil Transformasi Kasus HIV (a) Diagram kotak garis dan (b) Sebaran densitas

pencilan yang tersebar di kabupaten/kota seperti Kabupaten Nabire, Jayapura, Jakarta Pusat, Kabupaten Mimika, Kabupaten Keerom, Kota Denpasar, Kabupaten Jayawijaya, Kota Manado, dan Kabupaten Paniai.

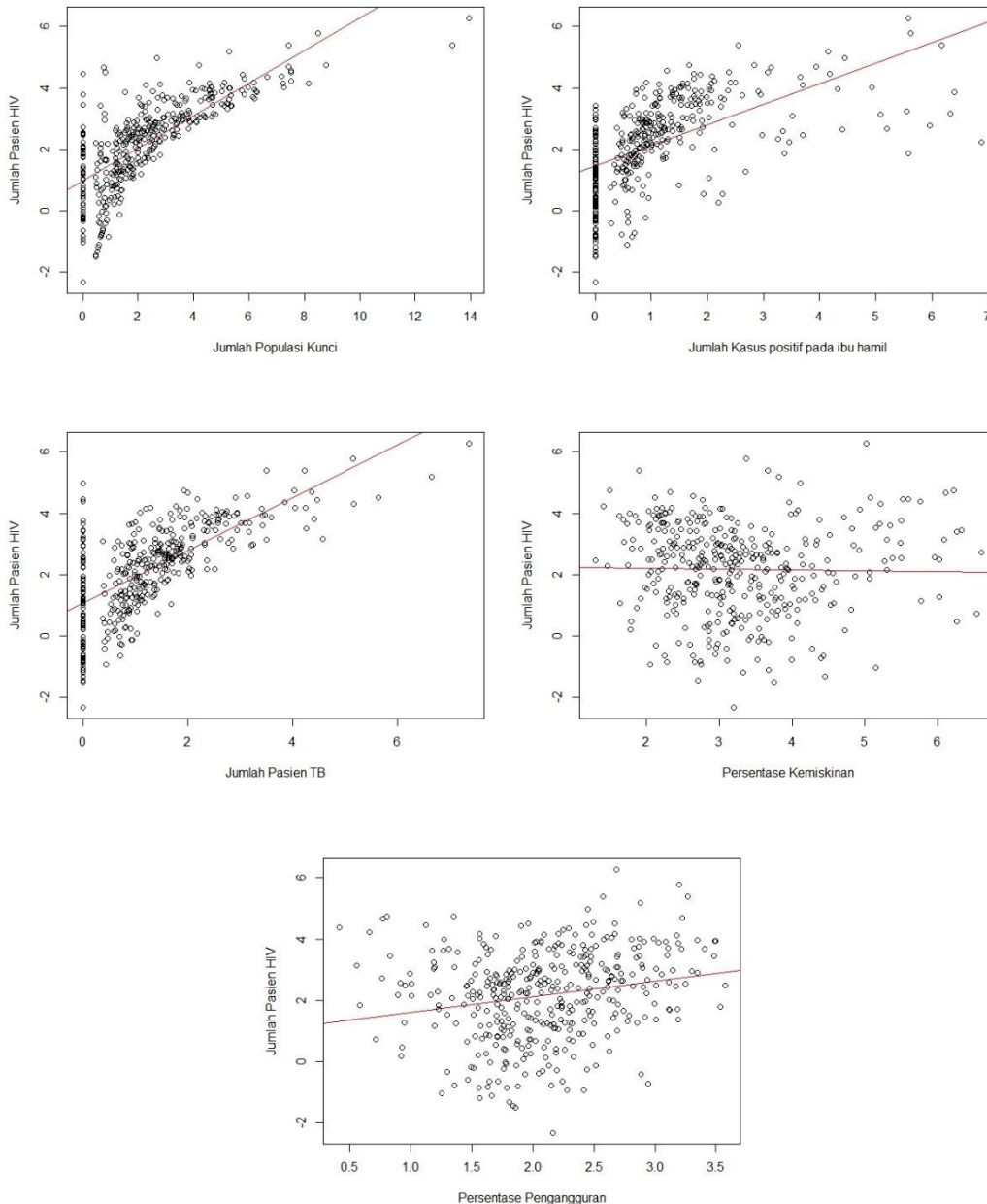
Pada **Error! Reference source not found.**, terlihat sebaran peubah respon setelah transformasi menggunakan logaritma natural. Amatan pencilan di luar batas atas terdapat di Kabupaten Nabire, sedangkan amatan pencilan di luar batas bawah terdapat di Kabupaten Langkat. Sebaran prediktor dalam penelitian ini disajikan dalam bentuk diagram kotak garis pada **Error! Reference source not found.** Seluruh prediktor telah ditransformasi menggunakan akar agar korelasi antar prediktor dan peubah respon lebih baik serta linier. Terlihat bahwa sebaran prediktor populasi kunci, kasus positif pada ibu



Gambar 3. Diagram Kotak Garis Prediktor Hasil Transformasi

hamil, pasien tuberkulosis, dan tingkat kemiskinan memiliki sebaran yang menjulur ke kanan dengan amatan pencilan yang terletak di ujung-ujung diagram kotak garis.

Hubungan antara peubah respon dengan prediktor yang telah dilakukan transformasi ditunjukkan pada **Error! R**



Gambar 4. Plot Sebaran Prediktor terhadap Kasus HIV

reference source not found. dan Error! Reference source not found.. Terlihat bahwa semakin tinggi prediktor pada populasi kunci, kasus positif pada ibu hamil, pasien tuberkulosis, dan tingkat pengangguran, maka semakin tinggi kasus HIV. Untuk peubah tingkat kemiskinan tidak menunjukkan adanya hubungan yang signifikan terhadap kasus HIV. Penjelasan ini di perkuat oleh penelitian Mufune (2015) yang menyatakan beberapa daerah mampu menurunkan tingkat infeksi HIV tanpa harus mengurangi tingkat kemiskinan.

Pengujian Multikolinieritas

Multikolinieritas antar prediktor merupakan salah satu persyaratan pada analisis regresi linier berganda. Berdasarkan pada

Tabel , setiap prediktor memiliki nilai VIF kecil yang berkisar antara 1,181 sampai dengan 1,657. Hal tersebut menunjukkan bahwa tidak terdapat multikolinieritas pada data.

Tabel 1. Nilai VIF Prediktor

Peubah	VIF
--------	-----

Populasi kunci	1,657
Kasus positif pada ibu hamil	1,356
Pasien tuberkulosis	1,541
Tingkat kemiskinan	1,181
Tingkat pengangguran	1,224

Pengujian Efek Spasial

Tabel 4. Statistik lima serangkai dugaan parameter lokal MGWR

Parameter	Min	Kuartil 1	Median	Kuartil 3	Maks
Intersep	-3,770	-0,463	0,385	1,381	4,150
Kasus positif pada ibu hamil	-0,054	0,118	0,218	0,301	0,813
Pasien tuberkulosis	-0,126	0,388	0,525	0,693	0,950
Tingkat kemiskinan	-0,730	-0,102	0,013	0,145	1,050
Tingkat pengangguran	-0,672	-0,110	0,010	0,238	1,456

Pemodelan regresi spasial diawali dengan melakukan uji dependensi spasial menggunakan uji RLM dan efek keragaman spasial dengan menggunakan uji BP. Berdasarkan Tabel , diperoleh hasil uji RLM nilai- p untuk $lag(\rho)$ sebesar $8,067 \times 10^{-9}$ yang berarti bahwa terdapat dependensi spasial pada $lag(\rho)$.

Tabel 2. Uji RLM

Parameter Statistik	Nilai- p
RLM (error)	0,765
RLM (lag)	$8,067 \times 10^{-9}$

Selanjutnya dilakukan uji BP diperoleh nilai- p sebesar $1,1 \times 10^{-4}$ yang menandakan bahwa terdapat efek keragaman spasial (heterogenitas spasial) pada data kasus HIV kabupaten/kota di Indonesia. Dengan demikian, data pada penelitian ini memiliki efek dependensi pada $lag(\rho)$ dan heterogenitas spasial. Salah satu analisis regresi yang dapat digunakan untuk mengatasi kedua hal tersebut adalah GWR dengan pendekatan SAR.

MGWR

Pemilihan fungsi pembobot yang paling sesuai menggunakan metode AIC. Tabel menyajikan AIC dari dua macam fungsi kernel dengan lebar jendela adaptif dan tetap. Fungsi pembobot *Kernel Bisquare* dengan lebar jendela adaptif memiliki AIC paling kecil.

Tabel 3. Perbandingan Fungsi Pembobot *Kernel Bisquare* Model GWR

Fungsi Pembobot	AIC
<i>Fixed Gaussian</i>	750,297
<i>Fixed Bisquare</i>	765,660
<i>Adaptive Gaussian</i>	736,442
<i>Adaptive Bisquare</i>	730,809

Berdasarkan hasil uji Wald parameter untuk populasi kunci berpengaruh signifikan terhadap seluruh kabupaten/kota yang menjadi lokasi amatan, seperti terlihat pada Gambar 5. Oleh karena itu, populasi kunci digunakan sebagai prediktor yang bersifat global, sedangkan kasus positif pada ibu hamil, pasien tuberkulosis, tingkat kemiskinan, dan tingkat pengangguran digunakan sebagai prediktor yang bersifat lokal, sehingga menyebabkan nilai pendugaan parameter yang berbeda untuk setiap kabupaten/kota. Brunson, Fotheringham, dan Charlton (1996) mengembangkan model GWR dengan



Gambar 5. Daerah signifikan peubah populasi kunci model GWR

menambahkan parameter global dan dikenal dengan MGWR.

Berdasarkan hasil GWR diketahui prediktor peubah kunci bersifat global dan prediktor lainnya bersifat lokal, sehingga dilanjutkan dengan MGWR. Pada Tabel 4 tertera dugaan parameter lokal MGWR dengan fungsi pembobot *Kernel Bisquare*. Pada Tabel 4 dapat dilihat bahwa dugaan parameter tingkat pengangguran memiliki *range* yang terbesar dibandingkan prediktor

lainnya, sedangkan dugaan parameter kasus positif pada ibu hamil memiliki *range* terkecil.

MGWR–SAR

Metode MGWR merupakan metode yang efektif untuk mengatasi data spasial yang memiliki keragaman antar lokasi, namun metode ini mengabaikan efek dependensi spasial. Sedangkan data penelitian ini memiliki efek dependensi spasial pada *lag* (ρ). Oleh karena itu, untuk menangani efek heterogenitas dan dependensi spasial pada data, maka digunakan MGWR–SAR.

Pada penelitian ini model MGWR–SAR yang digunakan ada dua macam yaitu MGWR–SAR ($0, k_c, k_v$) dan MGWR–SAR ($1, k_c, k_v$). Pada MGWR–SAR ($0, k_c, k_v$) mengasumsikan nilai *lag* (ρ) yang sama untuk setiap lokasi amatan, sedangkan MGWR–SAR ($1, k_c, k_v$) mengasumsikan nilai *lag* (ρ) yang berbeda untuk setiap lokasi amatan. Konstanta k_c adalah banyaknya parameter regresi global dan k_v adalah banyaknya parameter regresi lokal. Berikut hasil dari kedua macam model:

1. MGWR–SAR ($0, k_c, k_v$)

Model MGWR–SAR ($0, k_c, k_v$) dengan fungsi pembobot *Kernel Bisquare* menghasilkan dugaan parameter $\beta_v(u_i, v_i)$ yang berbeda di setiap lokasi amatan dan nilai *lag* (ρ) yang sama untuk setiap lokasi amatan yaitu sebesar -0,010, menunjukkan bahwa efek dependensi spasial lag negatif ini cukup kecil di seluruh wilayah yang diamati. Dugaan parameter $\beta_v(u_i, v_i)$ dirangkum dalam statistik lima serangkai pada **Error! Reference source not found.**

Hasil dugaan parameter MGWR dengan MGWR–SAR ($0, k_c, k_v$) sangat mirip. Dugaan parameter tingkat

pengangguran memiliki rentang terbesar, sedangkan prediktor kasus positif pada ibu hamil memiliki *range* terkecil (Gambar 6)

2. MGWR–SAR ($1, k_c, k_v$)

Model MGWR–SAR ($1, k_c, k_v$) dengan fungsi pembobot *Kernel Bisquare* menghasilkan dugaan parameter $\beta_v(u_i, v_i)$ dan nilai *lag* (ρ) yang berbeda di setiap lokasi amatan. Dugaan parameter $\beta_v(u_i, v_i)$ dan *lag* (ρ) dirangkum dalam statistik lima serangkai selengkapnya pada **Error! Reference source not found.**

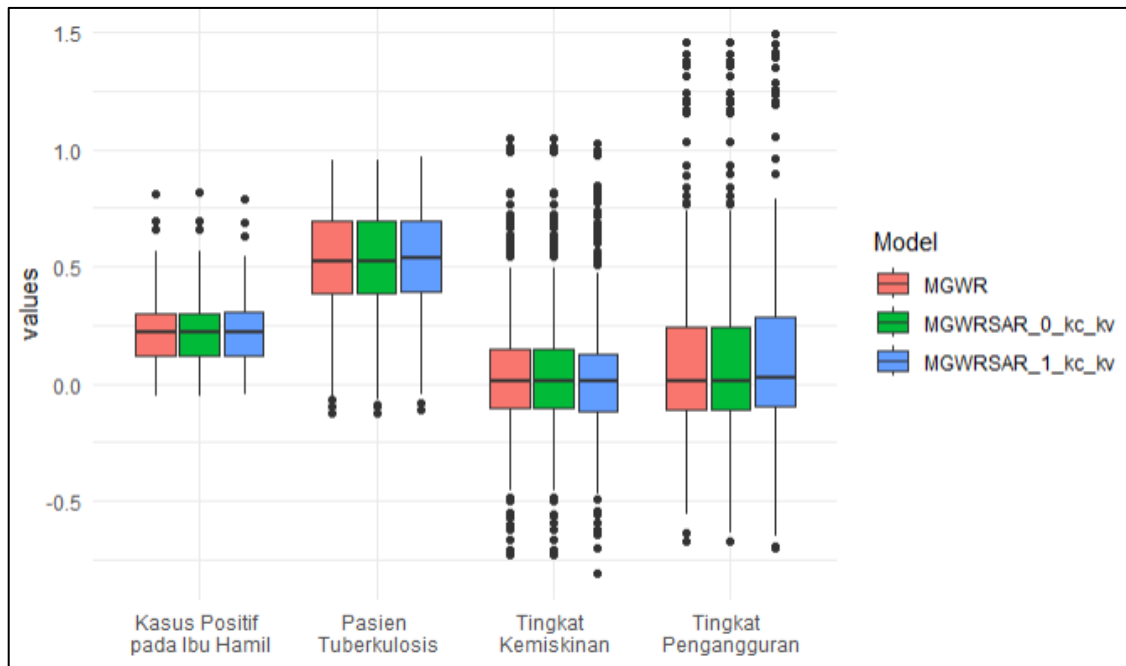
Pada Tabel 6 terlihat bahwa dugaan parameter untuk peubah kasus positif pada ibu hamil, pasien tuberkulosis, tingkat kemiskinan, tingkat pengangguran dan *lag* (ρ) memiliki *range* beragam dari nilai negatif sampai positif (Gambar 6). Dugaan parameter dari MGWR–SAR mirip dengan hasil MGWR, dugaan parameter tingkat pengangguran memiliki *range* terbesar dan pada kasus positif pada ibu hamil memiliki *range* terkecil.

Pemilihan Model terbaik

Model MGWR, MGWR–SAR ($0, k_c, k_v$), dan MGWR–SAR ($1, k_c, k_v$) dapat dibandingkan dengan melihat perbedaan nilai RMSE dari ketiga model tersebut. Pada Tabel , terlihat bahwa nilai RMSE terkecil adalah 0,54066 untuk model MGWR–SAR ($0, k_c, k_v$). Nilai RMSE antara MGWR–SAR ($0, k_c, k_v$) dan MGWR sangat kecil bedanya namun model MGWR–SAR ($0, k_c, k_v$) dipilih sebagai

Tabel 6. Statistik lima serangkai dugaan parameter lokal MGWR–SAR ($1, k_c, k_v$)

Parameter	Mini	Kuartil 1	Median	Kuartil 3	Maks
Intersep	-3,788	-0,445	0,401	1,405	4,162
Kasus positif pada ibu hamil	-0,044	0,117	0,217	0,305	0,789
Pasien tuberkulosis	-0,110	0,391	0,536	0,698	0,968
Tingkat kemiskinan	-0,810	-0,115	0,010	0,130	1,024
Tingkat pengangguran	-0,703	-0,095	0,024	0,287	1,494
<i>Lag</i> (ρ)	-0,990	-0,990	0,069	0,949	0,990



Gambar 6. Diagram kotak garis dugaan parameter lokal MGWR, MGWR–SAR $(0, k_c, k_v)$, MGWR–SAR $(1, k_c, k_v)$

model terbaik untuk menjelaskan pengaruh terhadap kasus HIV di seluruh kabupaten/kota Indonesia, karena data kasus HIV memiliki efek dependensi pada lag (ρ) dan heterogenitas spasial, model MGWR–SAR $(0, k_c, k_v)$ dapat mengatasi kedua hal tersebut.

Tabel 7. Perbandingan Model

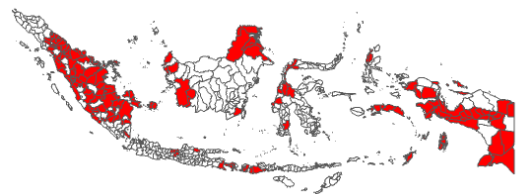
Model	RMSE
MGWR	0,54074
MGWR–SAR $(0, k_c, k_v)$	0,54066
MGWR–SAR $(1, k_c, k_v)$	0,75779

Pengujian signifikansi parameter diawali dengan parameter ρ dan parameter global dengan menggunakan uji Wald (**Error! Reference source not found.** Hasil uji menunjukkan parameter ρ dan parameter global berpengaruh signifikan terhadap kasus HIV di seluruh kabupaten/kota Indonesia.

Tabel 8. Hasil Uji Signifikansi Parameter

Parameter MGWR-SAR $(0, k_c, k_v)$	Koefisien	Nilai statistik uji Wald	Z_{tabel}	Keputusan
Lag (ρ)	-0,010	32,316	1,96	Menolak H_0
Global (populasi kunci)	0,313	13,547	1,96	Menolak H_0

Selanjutnya melakukan uji parsial untuk parameter lokal model MGWR-SAR $(0, k_c, k_v)$. Untuk meringkas hasil uji disajikan dalam peta. **Error! Reference source not found.** 7 menunjukkan kabupaten/kota dengan kasus positif pada ibu hamil signifikan (warna merah) terdapat pada hampir seluruh Provinsi di Sumatera dan di Provinsi Papua Barat dan Papua, untuk pulau Kalimantan pada Provinsi Kalimantan Utara dan Kalimantan Barat untuk Pulau Jawa Nusa Tenggara Barat dan



Gambar 7. Daerah signifikan peubah kasus positif pada ibu hamil model MGWR-SAR $(0, k_c, k_v)$

Bali, untuk Pulau Sulawesi pada Sulawesi Tengah.

Gambar 8 menunjukkan kabupaten/kota yang mempunyai peubah pasien tuberkulosis signifikan (warna ungu) mendominasi bagian Barat yaitu menyebar di Pulau Sumatera, Jawa, Kalimantan dan sebagian Sulawesi. Banyaknya kabupaten/kota yang signifikan pada peubah ini paling banyak dibandingkan prediktor lainnya.



Gambar 8. Daerah signifikan peubah pasien tuberkulosis model MGWR-SAR $(0, k_c, k_v)$

Error! Reference source not found. m enunjukkan bahwa sebaran peubah signifikan (warna hijau) tingkat kemiskinan menyebar sebagian di Provinsi Sumatera Utara, Provinsi Kalimantan Timur, Provinsi NTT dan Provinsi Maluku.



Gambar 9. Daerah signifikan peubah tingkat kemiskinan model MGWR-SAR $(0, k_c, k_v)$

Error! Reference source not found. m enunjukkan sebaran peubah signifikan (warna biru) tingkat pengangguran menyebar di sebagian Provinsi Sumatera Utara, Provinsi Sumatera Barat, Provinsi



Gambar 10. Daerah signifikan peubah tingkat pengangguran model MGWR-SAR $(0, k_c, k_v)$.

Riau, Provinsi Kalimantan Utara, Provinsi Kalimantan Timur, Provinsi NTT dan Provinsi Maluku.

Sebagai ilustrasi model MGWR-SAR $(0, k_c, k_v)$ pada Kabupaten Labuhan Batu, Sumatera Utara dengan semua prediktor lokal signifikan yaitu:

$$\hat{Y} = -0,010Wy + 0,313\text{populasi kunci} + 0,231\text{kasus positif ibu hamil} + 0,644\text{pasien TB} + 0,623\text{tingkat kemiskinan} + 0,419\text{tingkat pengangguran}$$

KESIMPULAN

MWGR-SAR sebagai penggabungan dari dua metode, MGWR dan SAR, dapat meningkatkan kemampuan model untuk menangkap keragaman spasial yang kompleks dan pola ketergantungan spasial. Penanganan efek dependensi dan heterogenitas spasial dalam MGWR-SAR mampu meningkatkan prediksi model pengaruh kasus HIV di seluruh kabupaten/kota Indonesia. MGWR-SAR $(0, k_c, k_v)$ efek dependensi pada *lag* (ρ) dan heterogenitas spasial yang dimiliki oleh data tersebut. Hanya saja penelitian ini menggunakan data tahun 2018 karena adanya keterbatasan pada sumber data.

Penerapan pada data HIV menunjukkan prediktor populasi kunci memiliki pengaruh signifikan secara global dan lainnya berpengaruh secara lokal. Prediktor Tuberkulosis merupakan koefisien lokal yang paling banyak signifikan dibandingkan dengan prediktor lokal lainnya. Untuk penelitian selanjutnya, diharapkan dapat menambahkan peubah-peubah baru yang dapat menunjukkan pengaruh kasus HIV di Indonesia.

DAFTAR PUSTAKA

- Anselin, L. (1988). Spatial Econometrics: Methods and Models, *Econometrica* [Preprint]. Kluwer Academic Publishers.

- <https://doi.org/10.1111/j.1468-0262.2004.00558.x>.
- Arbia, G. (2006). *Spatial Econometrics Statistical Foundations and Applications to Regional Convergence*. Berlin: Springer-Verlag.
- Brunsdon, C., Fotheringham, A.S. dan Charlton, M.E. (1996). 'Geographically weighted regression: a method for exploring spatial nonstationarity', *Geographical Analysis*, 28(4), pp. 281–298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>.
- Comber, L. dan Brunsdon, C. (2021). *Geographically Data Science & Spatial Data Analysis: An Introduction in R*. SAGE.
- Djuraidah, A. (2020). *Monograph Penerapan dan Pengembangan Regresi Spasial dengan Studi Kasus pada Kesehatan, Sosial, dan Ekonomi*. Bogor: IPB Press.
- El-Kautsar, M.B., Djuraidah, A. dan Angraini, Y. (2022). *Identifikasi faktor-faktor yang memengaruhi kasus HIV di Indonesia tahun 2018 menggunakan regresi terboboti geografis campuran*. Institut Pertanian Bogor.
- Fotheringham, A.S., Brunsdon, C. dan Charlton, M. (2002). *Geographically Weighted Regression: the analysis of spatially varying relationships*. John Wiley & Sons. Inc.
- Furková, A. (2021). Implementation of MGWR-SAR Models for Investigating a Local Particularity of European Regional Innovation Processes, *Central European Journal of Operations Research* [Preprint]. <https://doi.org/10.1007/s10100-021-00764-3>.
- Geniaux, G. dan Martinetti, D. (2018). A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models, *Regional Science and Urban Economics*, 72, pp. 74–85. <https://doi.org/10.1016/j.regsciurbeco.2017.04.001>.
- Kemendes RI (2019). *Profil Kesehatan Indonesia 2018, Health Statistics*. Jakarta.
- Kusnandar, D., Debatara, N.N. dan Fitriani, S. (2021). Pemodelan Sebaran Total Dissolved Solid Menggunakan Metode Mixed Geographically Weighted Regression, *Jurnal Aplikasi Statistika & Komputasi Statistik*, 13(1), pp. 9–16. <https://doi.org/10.34123/jurnalasks.v13i1.257>.
- Lu, B., Harris, P., Charlton, M., dan Brunsdon, C. (2014). The GWmodel R package: Further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-Spatial Information Science*, 17(2), 85–101. <https://doi.org/10.1080/10095020.2014.917453>
- Mufune, Pempelani. (2015). Poverty and HIV/AIDS in Africa: Specifying the connections, *Social Theory & Health*, 13(1), pp. 1-29.
- Olakunde, B.O., Parr, J.R., Adeyinka, D.A., Conserve, D.F., dan Duncan, D.T. (2021). Spatial analysis of HIV infection and the associated correlates among transgender persons in the United States, *Psychological and Socio-medical Aspects of AIDS/HIV*, 34(8). <https://doi.org/https://doi.org/10.1080/09540121.2021.1929817>.
- Shekhar, S., Li, Y., Ali, R.Y., Eftelioglu, E., Tang, X., dan Jiang, Z. (2018). Spatial and Spatiotemporal Data Mining, in B. Huang (ed.) *Comprehensive Geographic Information Systems*. Elsevier, pp. 264–286. <https://doi.org/https://doi.org/10.1016/B978-0-12-409548-9.09594-4>.

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

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

Bagian naskah berisi:

Judul. Judul tidak melebihi 12 kata dalam Bahasa Indonesia.

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

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

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

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

Metodologi terdiri atas:

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

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

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

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

Publikasi Buku

1. Penulis satu orang
Enders, Walter. 2010. *Applied Econometric Time Series, Third Edition*. New Jersey: Wiley.
2. Penulis dua orang
Pyndick, Robert. S. dan Rubinfeld, Daniel L. 2009. *Microeconomics, Seventh Edition*. New Jersey: Pearson Education.
3. Penulis tiga orang
Fotheringham, A. S., 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 September, 2014).

Buku yang ditulis oleh lembaga atau organisasi

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

Kertas kerja (working papers)

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

Makalah yang direpresentasikan

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

Karya yang tidak dipublikasikan

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

Artikel di koran, majalah, dan periodik sejenis

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