

JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 14, NOMOR 1, JUNI 2022 ISSN 2086 – 4132

Pemodelan Indeks Pembangunan Kesehatan Masyarakat Provinsi Jawa Timur Dengan Metode *Geographically Weighted Logistic Regression*

QURAINI SEPTY WARDHANI, SRI SULISTIOWATI HANDAJANI, dan IRWAN SUSANTO

Analisis Perbandingan Performa Metode Ensemble Dalam Menangani *Imbalanced Multi-Class Classification*

QORRY MEIDIANINGSIH dan DEVI EKA WARDANI MEGANINGTYAS

Klasifikasi Karakteristik Kemiskinan Di Provinsi Bengkulu Tahun 2020 Menggunakan Metode Pohon Klasifikasi Gabungan

WINALIA AGWIL, DIAN AGUSTINA, HERLIN FRANSISKA, dan NURUL HIDAYATI

Pemodelan Prevalensi Penyakit Kusta Di Jawa Timur Dengan Pendekatan *Geographically Weighted Regression*

CHRISTOPHER ANDREAS, HORIDAH HORIDAH, RIZZA SULISTIANA, DHYANA VENOSIA, dan NUR CHAMIDAH

Pandemi Covid-19 Dan *Turnover* Ke Pekerjaan Informal: Analisis Data Google Trends

ARI PURWANTO SARWO PRASOJO

Determinan Keluarnya Perempuan Dari Pasar Tenaga Kerja Pada Masa Pandemi Covid-19

TASMILAH TASMILAH



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.

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

Alamat Redaksi:

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

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

PENGANTAR REDAKSI

Puji syukur kehadiran Allah, Tuhan Yang Maha Esa, “Jurnal Aplikasi Statistika dan Komputasi Statistik” Volume 14, Nomor 1, Juni 2022 dapat diterbitkan. Jurnal ilmiah ini dapat terwujud atas partisipasi semua pihak, penulis internal dilingkungan Politeknik Statistika STIS maupun penulis eksternal, serta mitra bestari.

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

Jakarta, Juni 2022

Ketua Dewan Redaksi,

Dr. Nasrudin

JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 14, NOMOR 1, JUNI 2022

DAFTAR ISI

<u>Pengantar Redaksi</u>	iii
<u>Daftar Isi</u>	iv
<u>Abstrak</u>	v-xii
Pemodelan Indeks Pembangunan Kesehatan Masyarakat Provinsi Jawa Timur Dengan Metode <i>Geographically Weighted Logistic Regression</i> <u>Quraini Septy Wardhani, Sri Sulistijowati Handajani, dan Irwan Susanto</u>	1-12
Analisis Perbandingan Performa Metode Ensemble Dalam Menangani <i>Imbalanced Multi-Class Classification</i> <u>Qorry Meidianingsih dan Devi Eka Wardani Meganingtyas</u>	13-22
Klasifikasi Karakteristik Kemiskinan Di Provinsi Bengkulu Tahun 2020 Menggunakan Metode Pohon Klasifikasi Gabungan <u>Winalia Agwil, Dian Agustina, Herlin Fransiska, dan Nurul Hidayati</u>	23-32
Pemodelan Prevalensi Penyakit Kusta Di Jawa Timur Dengan Pendekatan <i>Geographically Weighted Regression</i> <u>Christopher Andreas, Horidah Horidah, Rizza Sulistiana, Dhyana Venosia, dan Nur Chamidah</u>	33-48
Pandemi Covid-19 Dan <i>Turnover</i> Ke Pekerjaan Informal: Analisis Data Google Trends <u>Ari Purwanto Sarwo Prasajo</u>	49-62
Determinan Keluarnya Perempuan Dari Pasar Tenaga Kerja Pada Masa Pandemi Covid-19 <u>Tasmilah Tasmilah</u>	63-76

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

DDC: 315.98

Kata kunci: IPKM, GWLR, *Adaptive Gaussian Kernel*

Quraini Septy Wardhani, Sri Sulistijowati Handajani, dan Irwan Susanto

Pemodelan Indeks Pembangunan Kesehatan Masyarakat Provinsi Jawa Timur Dengan Metode *Geographically Weighted Logistic Regression*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 14, Nomor 1, Juni 2022, hal 1 – 12

Abstrak

Geographically weighted logistic regression (GWLR) adalah metode yang dapat memodelkan data bersifat kategorik dengan mempertimbangkan efek spasial. Pada penelitian ini, metode GWLR digunakan untuk memodelkan indeks pembangunan kesehatan masyarakat (IPKM) di Provinsi Jawa Timur. IPKM merupakan indikator keberhasilan pemerintah dalam bidang kesehatan. Tingginya IPKM disuatu wilayah menunjukkan kualitas kesehatan yang baik di wilayah tersebut. Variabel yang digunakan adalah prevalensi balita *stunting*, hipertensi, pneumonia, persalinan ditangani nakes, pengguna KB dengan metode kontrasepsi jangka panjang (MKJP), dan penduduk dengan perilaku cuci tangan benar. Hasil penelitian ini menunjukkan bahwa pemodelan GWLR dengan pembobot *adaptive Gaussian Kernel* lebih baik daripada pembobot lain dengan faktor yang berpengaruh signifikan secara lokal adalah prevalensi balita *stunting* dan hipertensi. Nilai akurasi, *sensitivity*, dan *specificity* yang dihasilkan berturut-turut sebesar 97,4%, 100% dan 85,71%.

DDC: 315.98

Qorry Meidianingsih dan Devi Eka Wardani Meganingtyas

Analisis Perbandingan Performa Metode Ensemble Dalam Menangani *Imbalanced Multi-Class Classification*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 14, Nomor 1, Juni 2022, hal 13 – 22

Abstrak

Penelitian ini fokus pada membandingkan kinerja beberapa metode *ensemble* dalam mengatasi *imbalanced multi-class classification* dimana metode dekomposisi *one-versus-one* (OVO) diterapkan sehingga metode klasifikasi *Support Vector Machine* yang standar dapat digunakan. Data penelitian merupakan hasil bangkitan *software R* yang dirancang berdasarkan level persentase kelas minoritas, yaitu menjadi kategori *extreme*, *moderate*, dan *mild*. Selain itu, dirancang pula kombinasi jumlah kelas mayoritas dan minoritas yang mungkin terjadi sehingga terdapat sembilan jenis data simulasi. Hasil penelitian menunjukkan bahwa secara umum rata-rata ketepatan klasifikasi paling rendah diperoleh ketika data simulasi bersifat *moderate* dengan sebaran gugus data terbagi menjadi 2 kelas mayoritas dan 1 kelas minoritas. Metode *safe-level SMOTEBagging* memberikan performa yang paling baik bagi semua jenis data simulasi, terutama ketika gugus data bersifat ekstrim, yaitu ditunjukkan dengan nilai rata-rata akurasi yang diperoleh sebesar 98.60 persen. Kinerja metode klasifikasi SVM cukup baik dimana nilai rata-rata tingkat akurasi setiap kelas berkisar antara 67.80-98.60 persen.

Kata kunci: *imbalanced multi-class, ensemble method, bagging, underbagging, SMOTEBagging, safe-level SMOTE, support vector machine*

DDC: 315.98

Winalia Agwil, Dian Agustina, Herlin Fransiska, dan Nurul Hidayati

Klasifikasi Karakteristik Kemiskinan Di Provinsi Bengkulu Tahun 2020 Menggunakan Metode Pohon Klasifikasi Gabungan

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 14, Nomor 1, Juni 2022, hal 23 – 32

Abstrak

Kemiskinan adalah masalah yang mendesak diatasi baik pada tingkat nasional maupun global, yang diindikasikan sebagai salah satu prioritas utama dalam agenda dunia tentang Tujuan-tujuan Pembangunan Berkelanjutan atau *Sustainable Development Goals* (SDGs). Penanganan kemiskinan yang efektif akan membantu penyelesaian permasalahan dunia yang lain seperti permasalahan kelaparan, kesehatan, kesejahteraan, pendidikan, air bersih dan sanitasi. Studi ini bertujuan untuk mengklasifikasikan karakteristik kemiskinan rumahtangga di Propinsi Bengkulu, berdasarkan data hasil Survei Sosial Ekonomi Nasional (SUSENAS) 2020. Metode analisis data untuk mengidentifikasi karakteristik rumah tangga miskin dalam studi ini menggunakan *Classification and Regression Tree* (CART) – dengan menerapkan model *random forest* untuk mengatasi ketidakseimbangan data. Berdasarkan pemodelan tersebut, studi ini menemukan bahwa terdapat tiga variabel utama yang mencirikan rumah tangga miskin di propinsi Bengkulu, yaitu jumlah anggota rumah tangga, ijazah terakhir kepala rumah tangga dan luas lantai rumah. Temuan ini dapat digunakan sebagai dasar untuk identifikasi rumahtangga miskin, sehingga program-program bantuan diharapkan lebih tepat sasaran di masa mendatang.

Kata kunci: Kemiskinan, *Unbalanced dataset, CART, Random forest, Xgboost*

DDC: 315.98

Christopher Andreas, Horidah Horidah,
Rizza Sulistiana, Dhyana Venosia, dan Nur
Chamidah

Pemodelan Prevalensi Penyakit Kusta Di
Jawa Timur Dengan Pendekatan
Geographically Weighted Regression

Jurnal Aplikasi Statistika & Komputasi
Statistik, Volume 14, Nomor 1, Juni 2022,
hal 33 – 48

Abstrak

Indonesia merupakan penyumbang kasus baru penyakit kusta tiga terbesar lingkup dunia, dengan Jawa Timur sebagai provinsi yang berkontribusi tertinggi. Hal ini membuat pemerintah menetapkan target rencana aksi guna mengeliminasi kusta di Indonesia. Penyakit kusta tergolong penyakit menular sehingga penyebaran penyakit ini dipengaruhi faktor lingkungan geografis. Oleh sebab itu, digunakan metode *Geographically Weighted Regression* (GWR) karena dapat mengestimasi nilai parameter disetiap titik pengamatan melalui aspek spasial. Pada pemodelan prevalensi penyakit kusta di Jawa Timur, dengan persentase rumah tangga yang menggunakan fasilitas mandi, cuci, kakus bersama (X_1), persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak (X_2), dan jumlah penduduk miskin (X_3) memberi pengaruh signifikan terhadap prevalensi penyakit kusta di Jawa Timur, dengan sebaran kabupaten dan kota yang dipengaruhi secara signifikan masing – masing sebanyak 26,32%, 18,42%, dan 39,47%. Selain itu, hasil pemodelan GWR menunjukkan akurasi dan kebaikan model yang lebih baik daripada model regresi global yang ditinjau berdasarkan nilai *Symmetric Mean Absolute Percentage Error* (sMAPE) dan *Akaike Information Criterion* (AIC), yang masing-masing sebesar 12,14% dan -73,124. Pada penelitian ini, diketahui bahwa prevalensi penyakit kusta tertinggi di Jawa Timur terjadi di Kabupaten Sumenep dengan nilai

prevalensi sebesar 2,06. Untuk mencapai target eliminasi kusta yang ditetapkan pemerintah, maka Pemerintah Kabupaten Sumenep dapat menurunkan nilai persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak (X_2) hingga mencapai 12,01% dan jumlah penduduk miskin (X_3) hingga mencapai 13,215 ribu jiwa. Pemerintah Jawa Timur perlu fokus untuk menanggulangi variabel-variabel yang signifikan disetiap wilayah masing-masing kabupaten atau kota, supaya bisa menurunkan prevalensi kusta sampai target yang ditetapkan oleh pemerintah. Hasil penelitian dapat digunakan sebagai landasan dalam merumuskan rekomendasi kebijakan di setiap kabupaten dan kota di Jawa Timur guna mencapai target eliminasi kusta yang ditetapkan pemerintah.

Kata kunci: Analisis Spasial, Eliminasi Kusta, Metode GWR, Prevalensi Penyakit Kusta, Regresi Global

DDC: 315.98

Ari Purwanto Sarwo Prasajo

Pandemi Covid-19 Dan *Turnover* Ke Pekerjaan Informal: Analisis Data Google Trends

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 14, Nomor 1, Juni 2022, hal 49 – 62

Abstrak

Pandemi COVID-19 telah berdampak terhadap keberlangsungan pekerjaan bagi pekerja di Indonesia seperti penurunan pendapatan hingga pemutusan hubungan kerja (PHK). Menurunnya permintaan tenaga kerja memungkinkan pekerja terdampak PHK berpindah ke sektor informal. Dengan menggunakan data Google Trends, studi ini bertujuan untuk mengeksplorasi adanya sinyal pergeseran pekerjaan (*job turnover*) ke pekerjaan informal pada periode awal pandemi COVID-19, terutama saat pembatasan fisik dan normal baru di Indonesia. Empat kategori kata kunci: PHK, situs lowongan pekerjaan (loker), kurir, driver online, dan berjualan digunakan sebagai analisis terkait PHK dan pencarian kesempatan kerja yang bersifat informal. Hasil analisis tren dengan menggunakan regresi lokal (LOESS) dan metode difference-in-differences (DD) menemukan adanya sinyal *turnover* saat periode awal pandemi COVID-19, terutama setelah diberlakukannya pembatasan fisik. Sinyal *turnover* tersebut ditunjukkan oleh meningkatnya intensitas penelusuran terkait PHK yang diikuti dengan meningkatnya intensitas penelusuran terkait kesempatan pekerja atau usaha seperti cara berjualan online dan lowongan kurir.

Kata kunci: COVID-19, Google Trends, informal, normal baru, PHK

DDC: 315.98

Tasmilah Tasmilah

Determinan Keluarnya Perempuan Dari Pasar Tenaga Kerja Pada Masa Pandemi Covid-19

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 14, Nomor 1, Juni 2022, hal 63 – 76

Abstrak

Pandemi Covid-19 mengakibatkan peningkatan jumlah perempuan yang keluar dari pasar tenaga kerja. Dengan menggunakan data sakernas panel tahun 2019-2020, penelitian ini bertujuan untuk mengestimasi faktor penentu perempuan keluar dari pasar tenaga kerja dengan menggunakan regresi logistik. Hasil yang diperoleh bahwa keberadaan balita, status pekerjaan sebagai *self employed*, penggunaan internet, dan tempat bekerja di rumah akan meningkatkan peluang perempuan keluar dari pasar tenaga kerja pada masa pandemi Covid-19. Sedangkan status perempuan sebagai kepala rumah tangga, status pekerjaan sebagai buruh/karyawan, pendidikan tinggi, dan penggunaan teknologi digital akan menurunkan peluang perempuan keluar dari pasar tenaga kerja pada masa pandemi Covid-19.

Kata kunci: Pandemi Covid-19, angkatan kerja perempuan, pengangguran

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

DDC: 315.98

Quraini Septy Wardhani, Sri Sulistijowati Handajani, dan Irwan Susanto

Pemodelan Indeks Pembangunan Kesehatan Masyarakat Provinsi Jawa Timur Dengan Metode *Geographically Weighted Logistic Regression*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 14, Nomor 1, Juni 2022, hal 1 – 12

Abstract

Geographically weighted logistic regression (GWLR) is a method that can model categorical data by considering spatial effects. In this study, the GWLR method was used to model the public health development index (IPKM) in East Java Province. IPKM is an indicator of the government's success in the health sector. The high IPKM in an area shows the good quality of health in that region. The variables used are prevalence of stunting in children under five years, hypertension, pneumonia, childbirth handled by health workers, family planning users with long-term methods of contraception, and residents with proper hand washing behavior. The results of this study indicate that the GWLR modeling with the adaptive Gaussian Kernel weighting is better than the other weights with factors that have a significant local effect are the prevalence of stunting and hypertension under five. The resulting accuracy, sensitivity, and specificity values are 97.4%, 100% and 85.71%, respectively.

Keywords: PHDI, GWLR, Adaptive Gaussian Kernel.

DDC: 315.98

Qorry Meidianingsih dan Devi Eka Wardani Meganingtyas

Analisis Perbandingan Performa Metode Ensemble Dalam Menangani *Imbalanced Multi-Class Classification*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 14, Nomor 1, Juni 2022, hal 13 – 22

Abstract

This study focused on comparing the performance of several ensemble methods in handling imbalanced multi-class classification where the one-versus-one (OVO) decomposition method is applied so that the standard Support Vector Machine classification method can be used. The data were generated by R software which is designed based on the percentage level of the minority class, namely into “extreme”, “moderate”, and “mild” categories. In addition, a combination of the number of majority and minority classes that may occur is designed so that there are nine types of simulation data. The results showed that in general the lowest average classification accuracy was obtained when the simulation data was moderate with the distribution of the data set divided into 2 majority classes and 1 minority class. The safe-level SMOTEBagging method provides the best performance for all types of simulation data, especially when the data set is extreme, which is indicated by the average accuracy obtained of 98.60 percent. The performance of the SVM classification method is quite good where the average value of the accuracy level of each class ranges from 67.80-98.60 percent.

Keywords: imbalanced multi-class, ensemble method, bagging, underbagging, SMOTEBagging, safe-level SMOTE, support vector machine

DDC: 315.98

Winalia Agwil, Dian Agustina, Herlin Fransiska, dan Nurul Hidayati

Klasifikasi Karakteristik Kemiskinan Di Provinsi Bengkulu Tahun 2020 Menggunakan Metode Pohon Klasifikasi Gabungan

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 14, Nomor 1, Juni 2022, hal 23 – 32

Abstract

Poverty is a pressing issue at both country and global level, indicated as one of main priorities in the global agenda on Sustainable Development Goals (SDGs). Tackling poverty effectively will help solve other critical issues such as hunger, health, welfare, education, clean water, and sanitation. The aim of this study is to classify the main characteristics of poor households in Province of Bengkulu, using the data from the results of National Socioeconomic Survey (SUSENAS) 2020. Method of data analysis in identifying the poor household characteristics in the present study uses Classification and Regression Tree (CART) – specifically by applying a forest random model in attempt to adjust the unbalanced datasets. Main findings from this modelling application suggest that there are three variables that mainly characterized poor households in Bengkulu, namely the number of household members, the last education certificate of the household head, and the size of the house floor area. This finding is useful for a basis in identifying the poor households, thus intervention programs designed to assist the poor is expected to be more well-targeted in the future.

Keywords: Poverty, Unbalance datasets, CART, Random forest

DDC: 315.98

Christopher Andreas, Horidah Horidah, Rizza Sulistiana, Dhyana Venosia, dan Nur Chamidah

Pemodelan Prevalensi Penyakit Kusta Di Jawa Timur Dengan Pendekatan Geographically Weighted Regression

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 14, Nomor 1, Juni 2022, hal 33 – 48

Abstract

Indonesia is the largest contributor to new cases of leprosy in the world, with East Java as the province with the highest contribution. This has made the government set a target for an action plan to eliminate leprosy in Indonesia. Leprosy is classified as an infectious disease so that the spread of this disease is influenced by geographical environmental factors. Therefore, the Geographically Weighted Regression (GWR) method is used because it can estimate the parameter value at each observation point through the spatial aspect. In modeling the prevalence of leprosy in East Java, with the percentage of households using shared bathing, washing, and latrine facilities (X_1), the percentage of households that do not have access to sanitation feasible (X_2), and the number of poor people (X_3) had a significant influence on the prevalence of leprosy in East Java, with the distribution of districts and cities being significantly affected respectively by 26.32%, 18.42 %, and 39.47%. In addition, the results of the GWR modeling show that the accuracy and goodness of the model are better than the global regression model reviewed based on the Symmetric Mean Absolute Percentage Error (sMAPE) and Akaike Information Criterion (AIC) values, which are 12,14% and -73,124, respectively. In this study, it was found that the highest prevalence of leprosy in East Java occurred in Sumenep Regency with a prevalence value of 2,06. To achieve the leprosy elimination target set by

the government, the Sumenep Regency Government can reduce the percentage of households that do not have access to proper sanitation (X_2) to 12,01% and the number of poor people (X_3) to reach 13,215 thousand inhabitants. The East Java government needs to focus on tackling significant variables in each district or city, in order to reduce the prevalence of leprosy to the target set by the government. The results of the study can be used as a basis for formulating policy recommendations in every district and city in East Java to achieve the leprosy elimination target set by the government.

Keywords: Spatial Analysis, Elimination of Leprosy, GWR Method, Prevalence of Leprosy, Global Regression

DDC: 315.98

Ari Purwanto Sarwo Prasajo

Pandemi Covid-19 Dan *Turnover* Ke Pekerjaan Informal: Analisis Data Google Trends

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 14, Nomor 1, Juni 2022, hal 49 – 62

Abstract

The COVID-19 pandemic has impacted the labor market in Indonesia, such as decreasing wages and layoffs. The decline in worker demand may drive job turnover into the informal sector. Using Google Trends data, this study aims to explore the existence of informal job turnover signals during the early phase of the COVID-19 pandemic, particularly during physical restriction and new normal period in Indonesia. Four keyword categories: “PHK”, “situs loker”, “kurir”, “driver online”, and “berjualan” were used to analyze layoff and search the informal job opportunity. The trend analysis results using local regression (LOESS) and difference-in-differences (DD) methods found a signal of informal job turnover in the early phase of the COVID-19 pandemic, especially after physical restriction implementation. The job turnover signal was shown by increasing search intensity about “PHK” (layoff), followed by the search intensity of job opportunities such “berjualan online” (online selling) and “lowongan kurir” (courier jobs).

Keywords: COVID-19, Google Trends, informal, new normal, layoff

DDC: 315.98

Tasmilah Tasmilah

Determinan Keluarnya Perempuan Dari Pasar Tenaga Kerja Pada Masa Pandemi Covid-19

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 14, Nomor 1, Juni 2022, hal 63 – 76

Abstract

The Covid-19 pandemic has resulted in an increasing number of women leaving the labor market. Using Sakernas data 2019-2020, this study estimates the determinants of women leaving the labor market using logistic regression. The results obtained are that the presence of children under the age of five, employment status as self employed, use of the internet, and a place to work at home will increase women's chances of leaving the labor market during the Covid-19 pandemic. Meanwhile, the status of women as heads of households, employment status as workers/employees, higher education, and the use of digital technology will reduce women's chances of leaving the labor market during the Covid-19 pandemic.

Keywords: Covid-19 pandemi, female labor participant, Unemployment

PEMODELAN INDEKS PEMBANGUNAN KESEHATAN MASYARAKAT PROVINSI JAWA TIMUR DENGAN METODE GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION

Quraini Septy Wardhani¹, Sri Sulistijowati Handajani², Irwan Susanto³

^{1,2,3} Program Studi Statistika, FMIPA Universitas Sebelas Maret, Surakarta
e-mail: ²rr_ssh@staff.uns.ac.id

Received: November 11, 2021; Revised: February 8, 2022; Accepted: April 17, 2022; Published: June 30, 2022

Abstrak

Geographically weighted logistic regression (GWLR) adalah metode yang dapat memodelkan data bersifat kategorik dengan mempertimbangkan efek spasial. Pada penelitian ini, metode GWLR digunakan untuk memodelkan indeks pembangunan kesehatan masyarakat (IPKM) di Provinsi Jawa Timur. IPKM merupakan indikator keberhasilan pemerintah dalam bidang kesehatan. Tingginya IPKM disuatu wilayah menunjukkan kualitas kesehatan yang baik di wilayah tersebut. Variabel yang digunakan adalah prevalensi balita *stunting*, hipertensi, pneumonia, persalinan ditangani nakes, pengguna KB dengan metode kontrasepsi jangka panjang (MKJP), dan penduduk dengan perilaku cuci tangan benar. Hasil penelitian ini menunjukkan bahwa pemodelan GWLR dengan pembobot *adaptive Gaussian Kernel* lebih baik daripada pembobot lain dengan faktor yang berpengaruh signifikan secara lokal adalah prevalensi balita *stunting* dan hipertensi. Nilai akurasi, *sensitivity*, dan *specificity* yang dihasilkan berturut-turut sebesar 97,4%, 100% dan 85,71%.

Kata kunci: IPKM, GWLR, *Adaptive Gaussian Kernel*.

Abstract

Geographically weighted logistic regression (GWLR) is a method that can model categorical data by considering spatial effects. In this study, the GWLR method was used to model the public health development index (IPKM) in East Java Province. IPKM is an indicator of the government's success in the health sector. The high IPKM in an area shows the good quality of health in that region. The variables used are prevalence of stunting in children under five years, hypertension, pneumonia, childbirth handled by health workers, family planning users with long-term methods of contraception, and residents with proper hand washing behavior. The results of this study indicate that the GWLR modeling with the adaptive Gaussian Kernel weighting is better than the other weights with factors that have a significant local effect are the prevalence of stunting and hypertension under five. The resulting accuracy, sensitivity, and specificity values are 97.4%, 100% and 85.71%, respectively.

Keywords: PHDI, GWLR, *Adaptive Gaussian Kernel*.

PENDAHULUAN

Indeks pembangunan kesehatan masyarakat (IPKM) merupakan alat ukur yang digunakan pemerintah untuk mengetahui kondisi kesehatan masyarakat di suatu wilayah. IPKM memuat 30 indikator yang dikelompokkan ke dalam 7 sub-indeks. Sub-indeks yang termuat dalam IPKM yaitu kesehatan balita, kesehatan reproduksi, pelayanan kesehatan, perilaku kesehatan, penyakit tidak menular, penyakit menular, dan kesehatan lingkungan (Kemenkes, 2019). Data yang tercantum dalam setiap indikator diperoleh dari tiga survei yaitu riset kesehatan dasar (Riskesdas), survei sosial ekonomi nasional (Susenas), dan potensi desa (Podes) (Kemenkes, 2011).

Pada tahun 2018, Jawa Timur merupakan salah satu provinsi di Indonesia yang lebih dari 30% kabupaten/kotanya mengalami penurunan peringkat IPKM dari IPKM tahun 2013. Sepanjang tahun 2018, jumlah kematian bayi yang terjadi adalah sebanyak 4.016 bayi, dengan 4.338 kematian balita serta 3.148 kematian neonatal (Kemenkes, 2019). Selain itu, beberapa kasus kesehatan juga terjadi di Jawa Timur pada 2018, mulai dari tingginya kasus demam berdarah dengue (DBD), hingga angka *stunting* yang lebih tinggi dari nasional (Kemenkes, 2019). Oleh karena itu, diperlukan analisis statistik untuk mengetahui faktor-faktor yang mempengaruhi IPKM di Provinsi Jawa Timur dengan mempertimbangkan faktor spasial. Salah satu metode yang dapat digunakan untuk memodelkan data IPKM adalah metode *Geographically Weighted Logistic Regression* (GWLR).

GWLR adalah metode non-parametrik untuk menganalisis data spasial yang merupakan bentuk lokal dari regresi logistik (Desriwendi, dkk, 2015). Regresi logistik merupakan teknik analisis multivariat yang bertujuan untuk menjelaskan hubungan antara variabel independen dengan variabel dependen yang bersifat kategorik (Hosmer dan Lemeshow, 2000). Data variabel dependen pada model GWLR diasumsikan berdistribusi Bernoulli

yang menghasilkan nilai koefisien regresi pada setiap lokasi berbeda-beda (Fotheringham, dkk, 2002).

Penelitian dengan menggunakan metode GWLR sebelumnya pernah dilakukan oleh Faturrahman (2016) untuk mengetahui faktor-faktor yang mempengaruhi IPKM di Provinsi Papua. Selain itu, Albuquerque (2016) melakukan penelitian terkait pemodelan penilaian kredit dengan menggunakan metode GWLR, serta Sabat (2017) yang menggunakan metode GWLR untuk memodelkan data angka kematian bayi. Penelitian terkait dengan IPKM sebelumnya pernah dilakukan oleh Octavany (2017) yang melakukan penelitian terkait faktor yang berpengaruh terhadap IPKM di Provinsi Jawa Timur dengan pendekatan regresi *spline*. Faturrahman (2019) juga pernah melakukan penelitian mengenai pemodelan IPKM kabupaten/kota di Pulau Kalimantan dengan menggunakan pendekatan regresi probit. Dibandingkan dengan metode regresi *spline* dan probit yang telah dilakukan pada penelitian sebelumnya, metode GWLR mampu menyajikan model regresi berdasarkan lokasi sehingga akan terbentuk model sejumlah n lokasi. Perbedaan model tiap lokasi memungkinkan peneliti untuk mengetahui lebih spesifik terkait dengan permasalahan yang lebih mendesak dan harus segera diatasi di masing-masing lokasi.

METODOLOGI

Regresi Logistik

Model regresi logistik merupakan bentuk khusus dalam model linier umum (*Generalized Linear Models* atau GLM) di mana variabel acak dependen bernilai kategorik (0 dan 1) dan mengikuti distribusi Bernoulli. Model logistik ganda dapat dituliskan sebagai berikut (Agresti, 2007):

$$\pi(x_i) = \frac{\exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ji})}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ji})}$$

di mana $\pi(x_i)$ adalah observasi variabel dependen ke- i dengan $i = 1, 2, 3, \dots, n$; β_0 adalah intersep model regresi logistik;

β_j adalah nilai parameter regresi variabel independen ke- k ; dan x_{ji} adalah nilai observasi variabel independen ke- j pada lokasi ke- i . Pendugaan parameter regresi logistik menggunakan metode *maximum likelihood estimation* (MLE) dengan persamaan logaritma *likelihood* yang terbentuk yaitu:

$$\ell(\beta) = \sum_{j=0}^p \left(\sum_{i=1}^n y_i x_{ji} \right) \beta_j - \sum_{i=1}^n \ln \left(1 + \exp \left(\sum_{j=0}^p \beta_j x_{ji} \right) \right)$$

Penyelesaian pendugaan parameter regresi logistik dari fungsi logaritma *likelihood* dapat dilakukan dengan melalui prosedur *iteratively reweighted least square* (IRLS) dari metode *Newthon Raphson* (Lestari, 2020).

Pembobot Spasial

Pada analisis spasial, diperlukan pembobot spasial pada masing-masing lokasi ke- i . Sebelum melakukan perhitungan pembobot spasial, perlu dilakukan perhitungan jarak *Euclidean*. Apabila lokasi ke- j terletak pada koordinat (u_j, v_j) maka akan diperoleh jarak *Euclidean* antara lokasi ke- i dan lokasi ke- j sebagai berikut (Yrigoyen, dkk, 2007).

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

Terdapat beberapa fungsi pembobot *Kernel* yang dapat digunakan untuk analisis spasial yaitu (Foringham, dkk, 2002):

1. Fixed Gaussian Kernel

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

2. Adaptive Gaussian Kernel

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

3. Adaptive Bisquare Kernel

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

dengan h adalah *bandwith* yang bernilai sama untuk setiap lokasi; h_i adalah *bandwith* pada lokasi ke- i ; dan d_{ij} adalah jarak *Euclidean* antara lokasi (u_i, v_i) ke (u_j, v_j) .

Dalam mencari pembobot pada masing-masing lokasi didasarkan pada jarak *Euclidean* (d_{ij}) dan *bandwidth* (h) yang dihasilkan pada masing-masing lokasi. Pemilihan ukuran *bandwidth* yang optimum menjadi salah satu hal yang penting karena akan mempengaruhi ketepatan hasil regresi (Fotheringham, *et al.*, 2002). Pada penelitian ini digunakan metode *Generalized Cross Validation* (GCV) untuk penentuan *bandwidth* optimal dan pemilihan model terbaik. GCV merupakan modifikasi dari *Cross Validation* (CV) yang didapat dengan meminimumkan fungsi CV (Eubank, dkk, 1988). Nilai *bandwidth* (h) yang optimal diperoleh dari *bandwidth* yang menghasilkan nilai GCV minimum. GCV dapat dihitung dengan rumus (Suyitno, dkk.,2016):

$$GCV = \frac{n \sum_{i=1}^n [\hat{\pi}(u_i, v_i) - \bar{Y}]^2}{(n - v)^2}$$

dengan $v = \text{tr}(\mathbf{S})$, di mana matriks \mathbf{S} merupakan matriks berukuran $(n \times n)$ yang dituliskan dengan rumus $\mathbf{S} = \mathbf{X}(\mathbf{X}^T \mathbf{W} \mathbf{V} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{V}$; \mathbf{W} adalah matriks diagonal pembobot spasial; dan \mathbf{V} adalah matriks diagonal yang memiliki elemen diagonal $\hat{\pi}_i [1 - \hat{\pi}_i]$.

Geographically Weighted Logistic Regression (GWLR)

Geographically Weighted Logistic Regression (GWLR) adalah metode non parametrik yang merupakan bentuk lokal dari regresi logistik di mana lokasi diperhatikan dan diasumsikan bahwa data dari variabel dependen berdistribusi Bernoulli yang digunakan untuk menganalisis data spasial (Desriwendi, dkk., 2015). Model GWLR dapat dituliskan sebagai berikut (Caraka, dkk, 2017).

$$\pi(x_i) = \frac{\exp(\sum_{j=0}^p \beta_j(u_i, v_i) x_{ji})}{1 + \exp(\sum_{j=0}^p \beta_j(u_i, v_i) x_{ji})}$$

$; i = 1, 2, \dots, n$

dengan $\pi(x_i)$ adalah nilai observasi variabel dependen ke- i ; $\beta_{j(u_i, v_i)}$ adalah vektor nilai parameter variabel independen ke- j pada lokasi ke- i ; (u_i, v_i) merupakan koordinat letak geografis (*longitude*, *latitude*) dari lokasi ke- i ; dan x_{ji} adalah nilai observasi variabel independen ke- j pada lokasi ke- i .

Pendugaan parameter model GWLR dilakukan dengan metode *maximum likelihood estimation* (MLE) dengan fungsi logaritma *likelihood* untuk model GWLR dengan pembobot spasial ditunjukkan pada rumus persamaan berikut (Aji, 2014).

$$\ell^*(\beta_{(u_i, v_i)}) = \sum_{j=0}^p \left(\sum_{i=1}^n y_i w_{i(u_i, v_i)} \right) \beta_{j(u_i, v_i)} x_{ji} - \sum_{i=1}^n w_{i(u_i, v_i)} \ln \left(1 + \exp \left(\sum_{j=0}^p \beta_{j(u_i, v_i)} x_{ji} \right) \right)$$

Terdapat dua pengujian parameter yang harus dilakukan untuk memodelkan data dengan metode GWLR, yaitu pengujian parameter secara serentak dan parsial. Pengujian parameter GWLR secara serentak bertujuan untuk mengetahui signifikan tidaknya pengaruh antara variabel independen dan variabel dependen secara bersamaan atau serentak. Hipotesis uji yang dapat digunakan untuk pengujian parameter GWLR secara serentak sebagai berikut (Atkinson, 2003).

H_0 : $\beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_p(u_i, v_i) = 0$ (Tidak ada pengaruh yang signifikan antara variabel independen terhadap variabel dependen secara serentak).

H_1 : Paling sedikit satu $\beta_j(u_i, v_i) \neq 0$; $j = 1, 2, \dots, p$ (Ada pengaruh yang signifikan antara variabel independen terhadap variabel dependen secara serentak).

H_0 ditolak jika nilai p -value $< \alpha$ atau jika nilai $G_2 > \chi^2_{\alpha, p}$ dengan rumus pengujian statistik yang digunakan adalah

$$G_2 = 2(\ell(\hat{\Omega}_{GWLR}) - \ell(\hat{\omega}_{GWLR}))$$

$$P\text{-value}_{\text{serentak(GWLR)}} = P(G_v > G_2)$$

di mana

$$\ell(\hat{\Omega}_{GWLR}) = \sum_{i=1}^n (y_i \ln \hat{\pi}_i(u_i, v_i) + (1 - y_i) \ln (1 - \hat{\pi}_i(u_i, v_i)))$$

$$\ell(\hat{\omega}) = \sum_i^n (n_{1i} \ln(n_{1i}) + n_{0i} \ln(n_{0i}) - n \ln(n))$$

Pengujian parameter GWLR yang selanjutnya adalah pengujian parameter secara parsial. Pengujian ini bertujuan untuk mengetahui variabel independen mana yang berpengaruh signifikan terhadap variabel dependen. Hipotesis pengujian yang digunakan sebagai berikut (Foringham, 2002).

H_0 : $\beta_j(u_i, v_i) = 0$ (Tidak ada pengaruh yang signifikan antara variabel independen terhadap variabel dependen).

H_1 : $\beta_j(u_i, v_i) \neq 0$ (Ada pengaruh yang signifikan antara variabel independen terhadap variabel dependen).

H_0 ditolak jika nilai p -value $< \alpha$ atau $|W_j| > Z_{\alpha/2}$ dengan rumus pengujian statistik yang digunakan adalah

$$W_j = \frac{\hat{\beta}_j(u_i, v_i)}{SE(\hat{\beta}_j(u_i, v_i))}$$

$$P\text{-value}_{\text{parsial(GWLR)}} = P(Z > |W_j|)$$

dengan

$$SE(\hat{\beta}_j(u_i, v_i)) = \sqrt{\text{var} \hat{\beta}_j(u_i, v_i)}$$

Pengujian kesamaan model regresi logistik dan model GWLR menggunakan perbandingan nilai *deviance* model Regresi Logistik dan model GWLR dengan hipotesis sebagai berikut (Faturrahman, 2016).

H_0 : $\beta_j(u_i, v_i) = \beta_j$; $i = 1, 2, \dots, n, j = 1, 2, \dots, p$ (Tidak ada perbedaan yang signifikan antara model regresi logistik dan model GWLR)

H_1 : paling sedikit satu $\beta_j(u_i, v_i) \neq \beta_j$; $i = 1, 2, \dots, n, j = 1, 2, \dots, p$ (Ada perbedaan yang signifikan antara model Regresi Logistik dan model GWLR)

H_0 ditolak jika $F_{hit} > F_{(\alpha, v_1, v_2)}$. Misalkan model regresi logistik *deviance* dinyatakan dengan $D(\hat{\beta})$ dengan derajat bebas v_1 dan model GWLR dinyatakan dengan $D(\hat{\beta}(u_i, v_i))$ dengan derajat bebas v_2 maka statistik uji yang digunakan yaitu:

$$F_{hit} = \frac{D(\hat{\beta})/v_1}{D(\hat{\beta}(u_i, v_i))/v_2}$$

$$P\text{-value}_{\text{kesamaan(GWLR)}} = P(F_v > F_{hit})$$

Pemilihan Model Terbaik

Pemilihan model terbaik bertujuan untuk mencari metode yang sesuai untuk memodelkan data. Pada penelitian ini, pemilihan model terbaik dilakukan dengan membandingkan nilai *Generalized Cross Validation* (GCV), *Aikake Information Criterion* (AIC), dan koefisien determinasi (R^2). AIC merupakan pengukuran untuk kualitas relatif dari model statistik dari data yang diberikan. Semakin kecil nilai AIC maka model akan semakin baik. Rumus AIC didefinisikan sebagai berikut (Suyitno, dkk., 2016).

$$AIC = 2n \log(\hat{\sigma}) + n \log(2\pi) + n \left(\frac{n + v}{n - v - 2} \right)$$

Koefisien determinasi (R^2) adalah nilai yang menjelaskan seberapa besar kemampuan variabel independen dalam menjelaskan variabel dependen. Apabila nilai R^2 mendekati 1 maka model semakin baik. Rumus dari *R-squared* McFadden's didefinisikan sebagai (Lestari, 2020)

$$R_{MF}^2 = 1 - \frac{\ell(\hat{\Omega}_{GWLR})}{\ell(\hat{\omega}_{GWLR})}$$

Variabel Penelitian

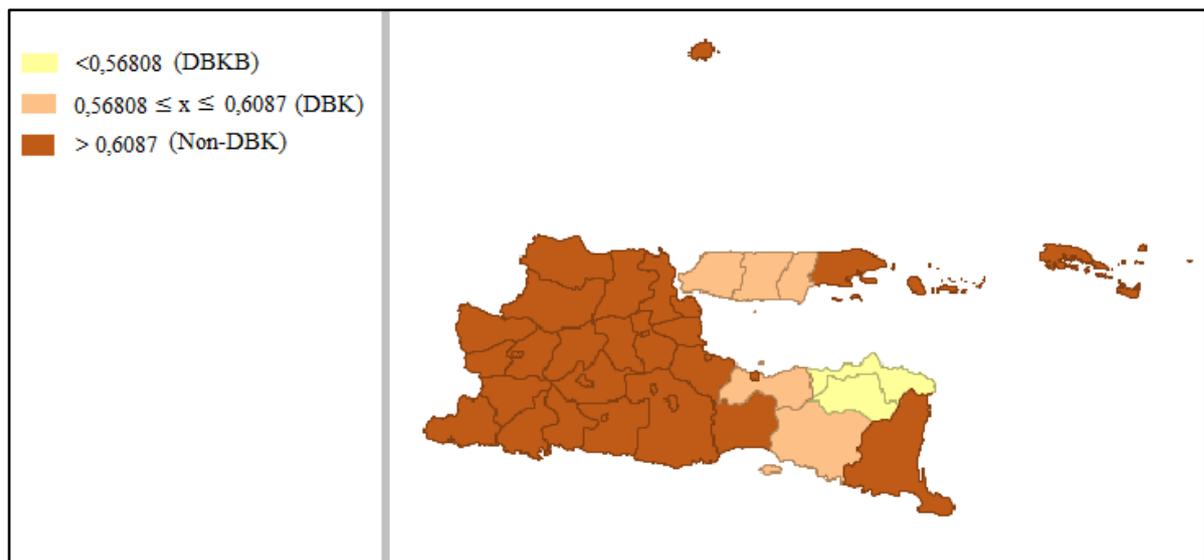
Data yang digunakan dalam penelitian ini merupakan data skunder yang diambil dari buku IPKM 2018 oleh Kementerian kesehatan Republik Indonesia yang diterbitkan pada tahun 2019. Terdapat

satu variabel dependen dan lima variabel independen pada penelitian ini. Variabel dependen yang digunakan adalah IPKM Provinsi Jawa Timur tahun 2018. Pada penelitian ini, IPKM dikelompokkan menjadi dua kategori, yaitu daerah tidak bermasalah kesehatan (non-DBK) dengan kategori 0 jika nilai IPKM $> 0,6087$ dan daerah bermasalah kesehatan (DBK) dengan kategori 1 jika nilai IPKM $< 0,6087$. Sedangkan untuk variabel independen yang digunakan, yaitu prevalensi balita *stunting* (X_1), prevalensi hipertensi (X_2), prevalensi pneumonia (X_3), persentase persalinan ditangani nakes (X_4), persentase pengguna KB MKJP (X_5), persentase proporsi penduduk dengan perilaku cuci tangan benar (X_6).

HASIL DAN PEMBAHASAN

Berikut merupakan statistik deskriptif dari data yang digunakan:

Pada peta tematik Gambar 1, menunjukkan bahwa terdapat dua daerah yang masuk dalam kelompok bermasalah kesehatan berat yaitu Kabupaten Bondowoso dan Kabupaten Situbondo. Selain itu, terdapat lima daerah yang masuk dalam kelompok daerah bermasalah kesehatan yaitu Kabupaten Jember, Probolinggo, Bangkalan, Sampang dan Pamekasan, sedangkan 31 daerah lainnya masuk dalam kelompok daerah non-bermasalah kesehatan.



Gambar 1. Klasifikasi Daerah Bermasalah Kesehatan Provinsi Jawa Timur Menurut Kemenkes

Pada penelitian ini hanya digunakan dua kelompok daerah yaitu daerah bermasalah kesehatan (DBK) dan daerah non-bermasalah kesehatan (non-DBK). Wilayah yang sebelumnya pada Gambar 4.1 masuk dalam kelompok daerah bermasalah kesehatan berat (DBKB) kemudian dimasukkan dalam kelompok daerah bermasalah kesehatan (DBK). Hal tersebut dikarenakan penelitian ini menggunakan metode regresi logistik dan *geographically weighted logistic regression* (GWLR) yang variabel dependennya bersifat kategori (0 dan 1) sehingga *cut-off* yang digunakan adalah rata-rata IPKM nasional tahun 2018.

Diagnosis Multikolinearitas

Berdasarkan Gambar 1. dapat diketahui bahwa terdapat 11 kabupaten/kota yang masuk dalam kategori daerah bermasalah kesehatan (DBK) dan 13 kabupaten/kota lainnya masuk dalam kategori non-DBK. Setelah dikelompokkan menjadi dua kategori, maka dapat dilakukan pemodelan dengan regresi logistik. Akan tetapi, sebelum melakukan pemodelan dengan regresi logistik terlebih dahulu dilakukan uji multikolinearitas untuk mendeteksi ada tidaknya korelasi antar variabel independen. Hasil yang diperoleh tertera pada Tabel 3.

Tabel 1 menunjukkan bahwa semua variabel independen yang digunakan memiliki nilai VIF < 10. yang artinya tidak terdapat multikolinearitas antar variabel independen.

Regresi Logistik

Terdapat dua uji parameter dalam memodelkan regresi logistik, yaitu uji parameter secara serentak dan parsial.

Tabel 3. Statistik Uji Wald

Variabel	Estimasi	Wald	P-value	Odds Ratio	Keputusan
Intercept	19,52	0,49957	0,617380	2,18x10 ⁸	Tidak Signifikan
X ₁	0,77233	1,93330	0,053196	2,164804	Signifikan
X ₂	-0,79265	-1,8069	0,070781	0,452644	Signifikan
X ₃	0,15279	0,10148	0,919170	1,165080	Tidak Signifikan
X ₄	-0,19706	-0,56783	0,570150	0,821141	Tidak Signifikan
X ₅	0,026207	0,11573	0,907870	1,026553	Tidak Signifikan
X ₆	-0,087736	-0,86285	0,388220	0,916003	Tidak Signifikan

Tabel 1. Nilai VIF Variabel Independen

Variabel	VIF
X ₁	3,697
X ₂	5,140
X ₃	1,183
X ₄	1,994
X ₅	1,369
X ₆	1,254

Tabel 2. Statistik Uji G

G ²	Df	P-value
25,811	6	0,000

Pengujian parameter secara serentak dilakukan dengan statistik uji G. Pada pengujian ini, H₀ akan ditolak jika $G^2 > \chi^2_{(6;0,1)} = 10,64$ atau ketika nilai $p\text{-value} < \alpha$. Statistik uji G untuk pengujian parameter secara serentak dapat dilihat pada Tabel 2.

Berdasarkan Tabel 2, dapat diketahui bahwa terdapat minimal satu variabel independen yang berpengaruh signifikan terhadap variabel respon secara serentak karena memiliki nilai statistik uji $G^2 = 25,811 > \chi^2_{(6;0,1)} = 10,64$ dan $p\text{-value} = 0,000 < \alpha = 0,1$ menunjukkan bahwa pada taraf signifikansi 0,1 terdapat minimal satu variabel independen yang berpengaruh signifikan, kemudian dilakukan pengujian parameter secara parsial dengan uji Wald.

Pada hasil estimasi parameter, dengan taraf signifikansi 0,1 akan diperoleh dua variabel independen yang berpengaruh signifikan terhadap variabel dependen yaitu X₁ dan X₂ (prevalensi balita *stunting* dan prevalensi penderita hipertensi). Berdasarkan Tabel 3 juga dapat dibentuk model regresi logistik seperti berikut:

$$\hat{\pi} = \frac{\exp(19,52 + 0,77233X_1 - 0,79265X_2 + 0,15279X_3 - 0,19706X_4 + 0,026207X_5 - 0,087736X_6)}{1 + \exp(19,5200 + 0,77233X_1 - 0,79265X_2 + 0,15279X_3 - 0,19706X_4 + 0,026207X_5 - 0,087736X_6)}$$

dengan model transformasi logitnya:

$$g(x) = 19,52 + 0,77233X_1 - 0,79265X_2 + 0,15279X_3 - 0,19706X_4 + 0,026207X_5 - 0,087736X_6$$

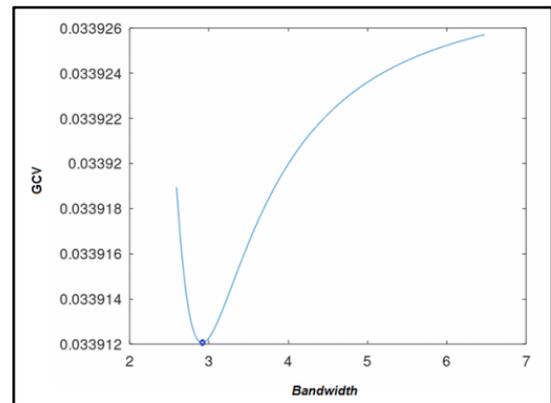
Interpretasi model logit di atas dapat dilakukan dengan cara mencari nilai *odds ratio* dari masing-masing parameter. Pada model di atas menjelaskan bahwa peningkatan prevalensi balita *stunting* sebesar 1% akan meningkatkan resiko Provinsi Jawa Timur masuk dalam kategori DBK sebesar 2,1648 kali. Apabila terjadi peningkatan prevalensi penderita hipertensi sebesar 1% akan meningkatkan resiko Provinsi Jawa Timur masuk dalam kategori DBK sebesar 0,4526 kali.

Pembobot Spasial

Penentuan *bandwidth* optimal sangat diperlukan dalam analisis regresi yang terboboti lokasi. *Bandwidth* optimum dipilih berdasarkan nilai *Generalized Cross Validation* (GCV) paling minimum. Sebelum menentukan *bandwidth* dari masing-masing lokasi, terlebih dahulu diperlukan untuk mencari jarak *Euclidean*.

Setelah jarak *Euclidean* terbentuk, maka dapat dibentuk *bandwidth* optimum dari masing-masing pembobot. Pada penelitian ini digunakan rentang *bandwidth*

optimum pada kisaran [bwb; bwa] di mana bwb(i) atau batas *bandwidth* bawah adalah nilai maksimal dari jarak *Euclidean* pada lokasi ke-i. Pemilihan bwb ini dikarenakan pada pembobot *adaptive Kernel* terdapat ketentuan bahwa apabila nilai $d_{ij} > h$ maka pembobot akan bernilai 0, sehingga untuk menghindari pembobot lokasi bernilai 0 maka batas bawah dari *bandwidth* adalah nilai maksimal dari jarak *Euclidean*. Sedangkan batas atas *bandwidth* (bwa)



Gambar 2. Grafik Bandwidth Optimal *Adaptive Bisquare* Lokasi Ke-19

yang digunakan dalam penelitian ini adalah 2,5 kali nilai maksimal jarak *Euclidean*. Dari interval *bandwidth* tersebut akan dipilih nilai *bandwidth* yang menghasilkan nilai *Generalized Cross Validation* (GCV) paling rendah. Grafik *bandwidth* yang terbentuk dengan pembobot *adaptive Bisquare Kernel* pada lokasi ke-19 (Kabupaten Madiun) tertera pada Gambar 2.

Lokasi ke-19 memiliki nilai *Euclidean* maksimal sebesar 2,5908 sehingga rentang *bandwidth* pada lokasi ke-19 atau Kabupaten Madiun adalah berkisar pada [2,5908; 6,477]. Pada Tabel 4

Tabel 4. Bandwidth optimal masing-masing lokasi

Lokasi	<i>Bandwidth</i> (<i>Fixed Gaussian Kernel</i>)	<i>Bandwidth</i> (<i>Adaptive Gaussian Kernel</i>)	<i>Bandwidth</i> (<i>Adaptive Bisquare Kernel</i>)
1	2,2926	2,7742	3,05833
2	2,2926	3,0714	3,5056
3	2,2926	5,7806	5,7715
4	2,2926	2,6935	2,6897
⋮	⋮	⋮	⋮
38	2,2926	2,2876	2,2926

menunjukkan *bandwidth* optimal masing-masing lokasi.

Setelah terbentuk matriks *bandwidth* optimum dari masing-masing pembobot, maka dilanjutkan dengan penentuan pembobot untuk masing-masing lokasi.

Pembobot *fixed Gaussian Kernel*:

$$w_{ij} = \begin{bmatrix} 1 & 0,99173 & \dots & 0,97763 \\ 0,99173 & 1 & \dots & 0,94361 \\ \vdots & \vdots & \ddots & \vdots \\ 0,97763 & 0,94361 & \dots & 1 \end{bmatrix}$$

Pembobot *adaptive Gaussian Kernel*:

$$w_{ij} = \begin{bmatrix} 1 & 0,99538 & \dots & 0,93834 \\ 0,99434 & 1 & \dots & 0,97789 \\ \vdots & \vdots & \ddots & \vdots \\ 0,98467 & 0,96818 & \dots & 1 \end{bmatrix}$$

Pembobot *adaptive Bisquare Kernel*:

$$w_{ij} = \begin{bmatrix} 1 & 0,98584 & \dots & 0,91156 \\ 0,98142 & 1 & \dots & 0,78132 \\ \vdots & \vdots & \ddots & \vdots \\ 0,9498 & 0,90317 & \dots & 1 \end{bmatrix}$$

Geographically Weighted Logistic Regression

Model GWLR akan terbentuk setelah pembobot untuk masing-masing lokasi ditemukan. Model akan terbentuk dengan nilai estimasi parameter yang berbeda untuk setiap lokasi. Sama halnya dengan regresi logistik global, terdapat tiga pengujian yang harus dilakukan sebelum membentuk model GWLR yaitu, pengujian parameter secara serentak, pengujian parameter parsial, dan uji kesamaan model. Pengujian parameter serentak dilakukan dengan tujuan untuk mengetahui apakah variabel independen secara bersama-sama berpengaruh signifikan terhadap variabel dependen dengan metode GWLR.

Hasil pengujian parameter secara serentak untuk masing-masing pembobot tertera pada Tabel 5.

Tabel 5 menunjukkan bahwa pada taraf signifikansi 10% terdapat pengaruh yang signifikan antara variabel independen terhadap variabel dependen secara serentak semua model GWLR baik dengan pembobot *fixed Gaussian*, *adaptive Gaussian*, maupun *adaptive Bisquare*.

Setelah dilakukan uji parameter GWLR secara serentak maka dilanjutkan dengan pengujian kesamaan model GWLR perlu dilakukan untuk melihat apakah terdapat perbedaan yang signifikan antara model GWLR dan regresi logistik global atau tidak. Pengujian dapat dilakukan dengan menggunakan uji *goodness of fit* di mana akan dilihat berdasarkan statistik uji F-nya.

Tabel 6 menunjukkan bahwa semua nilai F_{hitung} dari pembobot *fixed Gaussian*, *adaptive Gaussian*, maupun *adaptive Bisquare Kernel* memiliki nilai di bawah $F_{(0,1;32;32)} = 1,582$ maka dapat diperoleh kesimpulan untuk tidak menolak H_0 pada taraf signifikansi sebesar 10% yang artinya tidak terdapat perbedaan yang signifikan antara model regresi logistik global dengan model GWLR baik dengan pembobot *fixed Gaussian*, *adaptive Gaussian*, maupun *adaptive Bisquare*.

Sebelum melakukan pengujian parameter secara parsial, terlebih dahulu dilakukan pemilihan model GWLR terbaik. Pemilihan model ini dilakukan dengan membandingkan nilai GCV, AIC dan R^2 .

Tabel 5. Pengujian Signifikansi Parameter GWLR Secara Serentak

Model	G_2	$\chi^2_{(6;0,1)}$	P-value	Keputusan
GWLR (<i>fixed Gaussian</i>)	25,254		0,00141	H_0 ditolak
GWLR (<i>adaptive Gaussian</i>)	25,516	10,64	0,00127	H_0 ditolak
GWLR (<i>adaptive Bisquare</i>)	24,370		0,00668	H_0 ditolak

Tabel 6. Uji Kesamaan Model GWLR

Model	F_{hitung}	$F_{(0,1;32;32)}$	P-value	Keputusan
<i>Fixed Gaussian</i>	1,0122		0,4864	H_0 tidak ditolak
<i>Adaptive Gaussian</i>	1,0147	1,582	0,4837	H_0 tidak ditolak
<i>Adaptive Bisquare</i>	1,056		0,4393	H_0 tidak ditolak

Tabel 7. Perbandingan Model

Model GWLR	GCV	AIC	R ²
<i>Fixed Gaussian Kernel</i>	0,055803	34,930	0,7127
<i>Adaptive Gaussian Kernel</i>	0,055013	34,164	0,7161
<i>Adaptive Bisquare Kernel</i>	0,05601	37,757	0,7292

Tabel 7 menunjukkan bahwa terdapat kontradiksi antara nilai AIC dan R² pada model GWLR pembobot *adaptive Gaussian Kernel* serta *adaptive Bisquare Kernel*. Nilai AIC minimum dimiliki oleh model GWLR pembobot *adaptive Gaussian Kernel*, sedangkan nilai R² maksimum dimiliki oleh model GWLR pembobot *adaptive Bisquare Kernel*. Meskipun terdapat kontradiksi, akan tetapi metode terbaik dapat dipilih dengan melihat nilai GCV masing-masing model. Nilai GCV minimum dimiliki oleh model GWLR pembobot *adaptive Gaussian Kernel* sehingga metode terbaik untuk memodelkan data IPKM Provinsi Jawa Timur tahun 2018 adalah metode GWLR dengan pembobot *adaptive Gaussian Kernel* karena memenuhi dua kriteria pemilihan model. Kemampuan variabel independen dalam menjelaskan variabel dependen dengan model GWLR pembobot *adaptive Gaussian Kernel* adalah sebesar 71,61%.

Pengujian yang selanjutnya adalah pengujian parameter dari masing-masing lokasi untuk melihat variabel mana yang berpengaruh signifikan. Model terbaik yang telah dipilih adalah model GWLR pembobot *adaptive Gaussian Kernel*, maka pengujian parameter parsial dilakukan pada model GWLR pembobot *adaptive Gaussian Kernel*.

Rangkuman dari kelompok kabupaten/kota berdasarkan variabel signifikan dari metode GWLR dengan pembobot *adaptive Gaussian Kernel* untuk memudahkan dalam membaca peta tematik pada Gambar 2 tertera pada Tabel 8 berikut.

Metode GWLR mampu menghasilkan model yang berbeda di masing-masing kabupaten dan kota di Jawa Timur sehingga pada Provinsi Jawa Timur, metode ini akan menghasilkan sebanyak 38 model yang memiliki nilai estimasi parameter yang berbeda pada setiap lokasi. Model GWLR dengan pembobot *adaptive Gaussian Kernel* untuk lokasi ke-11 (Kabupaten Bondowoso) sebagai berikut.

$$\hat{\pi}(u_{11}, v_{11}) =$$

$$\frac{\exp(19,19 + 0,76744X_1 - 0,78653X_2 + 0,15533X_3 - 0,19422X_4 + 0,026812X_5 - 0,08714X_6)}{1 + \exp(19,19 + 0,76744X_1 - 0,78653X_2 + 0,15533X_3 - 0,19422X_4 + 0,026812X_5 - 0,08714X_6)}$$

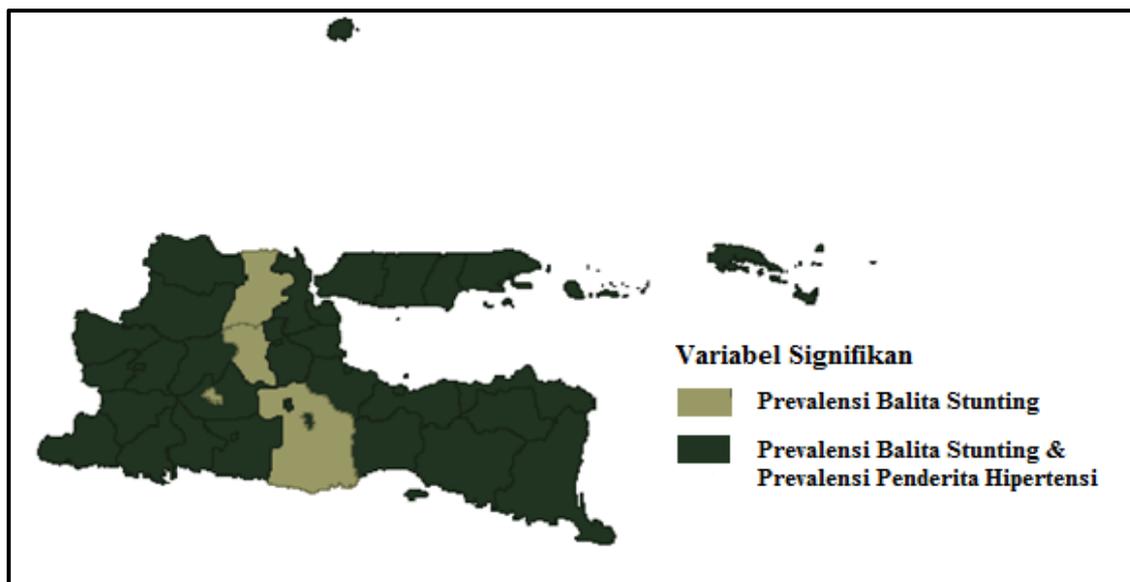
Transformasi logitnya:

$$g(x)_{(u_{11}, v_{11})} = 19,19 + 0,76744X_1 - 0,78653X_2 + 0,15533X_3 - 0,19422X_4 + 0,026812X_5 - 0,08714X_6$$

Interpretasi model untuk masing-masing lokasi dapat dilakukan dengan menghitung nilai *odds ratio* dari setiap parameter. Nilai *odds ratio* inilah yang akan memudahkan dalam memahami model yang terbentuk. Perhitungan *odds ratio*

Tabel 8. Variabel Yang Signifikan pada Model GWLR dengan Pembobot Adaptive Gaussian Kernel

Signifikan	Kabupaten/Kota
X ₁ (prevalensi balita <i>stunting</i>)	Kediri, Malang, Jombang, dan Lamongan.
X ₁ (prevalensi balita <i>stunting</i>), X ₂ (prevalensi penderita hipertensi)	Pacitan, Banyuwangi, Bondowoso, Probolinggo, Pasuruan, Gresik, Kota Blitar, Jember, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Madiun, Magetan, Bojonegoro, Tuban, Bangkalan, Sampang, Pamekasan, Situbondo, Sidoarjo, Mojokerto, Nganjuk, Ngawi, Sumenep, Kota Madiun, Kota Kediri, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Malang, Kota Surabaya, dan Kota Batu.



Gambar 3. Peta Tematik Berdasarkan Variabel Signifikan GWLR Adaptive Bisquare Kernel

dilakukan dengan mencari nilai exponential dari setiap parameter (β) sehingga interpretasi model GWLR dengan pembobot *adaptive Gaussian Kernel* untuk lokasi ke-11 (Kabupaten Bondowoso) yaitu apabila terjadi peningkatan prevalensi balita *stunting* sebesar 1% akan meningkatkan resiko Kabupaten Bondowoso masuk dalam kategori DBK sebesar 2,15489 kali serta adanya peningkatan prevalensi penderita hipertensi sebesar 1% akan meningkatkan resiko Kabupaten Bondowoso masuk dalam kategori DBK sebesar 0,45542 kali.

Hasil estimasi dari GWLR tidak hanya mampu menampilkan resiko suatu wilayah masuk dalam kategori daerah bermasalah kesehatan (DBK) berdasarkan masing-masing variabel independennya,

akan tetapi metode ini juga mampu menunjukkan probabilitas suatu wilayah masuk dalam kategori DBK secara keseluruhan. Nilai probabilitas ini dilihat berdasarkan besaran nilai $\hat{\pi}$ dari masing-masing kabupaten/kota. Nilai $\hat{\pi}$ berada dalam rentang [0,1]. Semakin besar nilai $\hat{\pi}$ maka probabilitas wilayah tersebut akan masuk dalam kategori daerah bermasalah kesehatan semakin tinggi. Matriks probabilitas untuk masing-masing wilayah di Provinsi Jawa Timur seperti table 9.

Nilai $\hat{\pi}$ untuk lokasi ke-11 menunjukkan bahwa probabilitas Kabupaten Bondowoso masuk dalam kategori daerah bermasalah kesehatan adalah sebesar 0,87798 yang artinya Kabupaten Bondowoso beresiko tinggi

Tabel 9. Variabel Yang Signifikan pada Model GWLR dengan Pembobot Adaptive Gaussian Kernel

Lokasi	$\hat{\pi}$	Lokasi	$\hat{\pi}$	Lokasi	$\hat{\pi}$
1	$3,71 \times 10^{-6}$	14	0,20725	27	0,99164
2	0,001068	15	0,000438	28	0,98199
3	0,19862	16	$7,85 \times 10^{-5}$	29	0,22875
4	$2,18 \times 10^{-5}$	17	$6,00 \times 10^{-5}$	30	0,013221
5	$8,78 \times 10^{-5}$	18	$2,53 \times 10^6$	31	$2,62 \times 10^{-6}$
6	0,001414	19	$2,28 \times 10^{-5}$	32	$9,24 \times 10^{-5}$
7	$5,31 \times 10^{-5}$	20	$9,29 \times 10^{-6}$	33	0,038221
8	0,043799	21	0,41812	34	0,34006
9	0,061822	22	0,0028	35	$2,65 \times 10^{-7}$
10	0,000295	23	0,024739	36	$7,85 \times 10^{-9}$
11	0,87798	24	0,039521	37	0,001573
12	0,7421	25	$5,29 \times 10^{-5}$	38	$7,52 \times 10^{-8}$
13	0,98018	26	0,85876		

untuk masuk dalam kategori daerah bermasalah kesehatan. Nilai $\hat{\pi}$ lokasi ke-38 menunjukkan bahwa probabilitas Kota Batu masuk dalam kategori daerah bermasalah kesehatan sangat rendah sehingga Kabupaten Bondowoso jauh lebih beresiko daripada Kota Batu.

Perhitungan akurasi, *sensitivity*, dan *specificity* pada model GWLR pembobot *adaptive Gaussian Kernel* sebagai berikut.

$$\text{Akurasi} = \frac{(31 + 6)}{31} = 0,9736$$

$$\text{Sensitivity} = \frac{31}{31} = 1$$

$$\text{Specificity} = \frac{6}{7} = 0,8571$$

Nilai akurasi model GWLR pembobot *adaptive Gaussian Kernel* sebesar 0,9736 menunjukkan bahwa kemampuan model dalam memprediksi secara benar adalah 97,36%. Nilai *sensitivity* sebesar 1 menunjukkan bahwa kemampuan model dalam memprediksi daerah non-bermasalah kesehatan dengan benar sebesar 100%. Sedangkan nilai *specificity* sebesar 0,8571 menunjukkan kemampuan model memprediksi daerah bermasalah kesehatan dengan benar sebesar 85,71%.

KESIMPULAN DAN SARAN

Meskipun tidak terdapat perbedaan yang signifikan antara model GWLR dengan regresi logistik global, metode GWLR merupakan salah satu metode yang dapat digunakan untuk memodelkan IPKM karena dengan menggunakan metode ini dapat diketahui faktor yang mempengaruhi suatu daerah masuk dalam kategori daerah bermasalah kesehatan (DBK) secara signifikan. Metode GWLR terbaik yang dapat digunakan untuk melakukan pemodelan data IPKM Provinsi Jawa Timur tahun 2018 adalah metode GWLR dengan pembobot *adaptive Gaussian Kernel* karena memiliki nilai GCV dan AIC paling kecil dibandingkan dua pembobot lainnya di mana nilai GCV-nya sebesar 0,055013 dan AIC sebesar 34,164. Pemodelan dengan GWLR dengan pembobot *adaptive Gaussian Kernel* membentuk satu variabel yang berpengaruh signifikan secara global

dan satu variabel yang berpengaruh secara lokal. Variabel yang berpengaruh secara global adalah X_1 (prevalensi balita *stunting*) sedangkan variabel yang berpengaruh signifikan secara lokal adalah X_2 (prevalensi penderita hipertensi).

Pada penelitian ini digunakan 6 variabel independen yang mewakili 7 indikator dari IPKM. Pada penelitian selanjutnya, peneliti menyarankan untuk menggunakan 7 variabel independen yang mana mewakili masing-masing satu variabel dari setiap indikator dan wilayah penelitian dapat dipersempit menjadi kecamatan atau bahkan desa karena semakin kecil satuan wilayah maka penelitian akan semakin tepat sasaran. Selain itu, penelitian menggunakan metode GCV untuk mencari *bandwith* optimal, pada penelitian selanjutnya dapat menggunakan metode pencarian *bandwidth* optimum lain seperti *golden section search*, CV, maupun AICc yang diharapkan dapat membentuk model GWLR yang berbeda signifikan dengan regresi logistik global.

DAFTAR PUSTAKA

- Aji, Caraka. A. W. 2014. Analisis Faktor-Faktor yang Mempengaruhi Laju Pertumbuhan Penduduk Kota Semarang Tahun 2011 Menggunakan Geographically Weighted Logistic Regression. *Skripsi*. Semarang.
- Albuquerque, P., Median, F. dan Silva, A. 2016. Geographically Weighted Logistic Regression Applied to Credit Scoring Models. *Revista Contabilidade and Financas*. Vol. 28 (73), 93-112.
- Atkinson, P. M., German, S. E., Sear, D. A., & Clark, M. J. (2003). Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. *Geographical Analysis*, Vol. 35(1), 58-82.
- Caraka, Rezzy. E. dan Yasin, Hasbi. 2017. *Geographically Weighted Regression (GWR)*. Yogyakarta: Mobius.
- Desriwendi, Hoyyi, A. dan Wuryandari, T. 2015. Pemodelan *Geographically Weighted Logistic Regression (GWLR)* dengan Fungsi Pembobot Fixed

- Gaussian Kernel dan Adaptive Gaussian Kernel. *Jurnal Gaussian*. Vol. 4 (2), 193-204.
- Eubank, Randall. 1988. Spline Smoothing and Nonparametric Regression. *Journal of the American Statistical Association*. Vol. 96 (456), 1523-1524.
- Faturrahman, M. 2019. Pemodelan Indeks Pembangunan Kesehatan Masyarakat Kabupaten/Kota Di Pulau Kalimantan Menggunakan Pendekatan Regresi Probit. *Jurnal Varian*. Vol. 2 (2), 47-54.
- Faturrahman, M., Purhadi, Sutikno, dan Ratnasari, V. 2016. Pemodelan Geographically Weighted Logistic Regression pada Indeks Pembangunan Kesehatan Masyarakat di Provinsi Papua. *Prosiding Seminar Nasional MIPA*. 34-42.
- Fotheringham, A. S., Lu, B., Charlton, M. dan Harris, P. 2002. Geographically Weighted Regression with a non-Euclidean Distance Metric: a Case Study Using Hedonic House Price Data. *International Journal of Geographical Information Science*. <http://dx.doi.org/10.1080/13658816.2013.865739> (Diakses 25 September, 2020).
- Hosmer, David. W. dan Lemeshow, Stanley. 1989. *Applied Logistic Regression*. Edisi Kedua. New York: A Willey-Interscience Publication.
- Kemenkes. 2011. Buku Saku Penanggulangan Daerah Bermasalah Kesehatan. Jakarta: Kemenkes.
- Kemenkes. 2019. *IPKM 2018*. Jakarta: Lembaga Penerbit Badan Penelitian dan Pengembangan Kesehatan (LPB).
- Kemenkes. 2019. Profil Kesehatan Provinsi Jawa Timur Tahun 2018. Surabaya: Kemenkes.
- Lestari, Vivi. D. 2020. Model Geographically Weighted Logistic Regression dengan Fungsi Pembobot Adaptive Tricube. *Skripsi*. Samarinda.
- Octavany, M. A., Budiantara dan Ratnasari. 2017. Pemodelan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Kesehatan Masyarakat Provinsi Jawa Timur Menggunakan Pendekatan Regresi Semiparametrik Spline. *Jurnal Sains dan Seni ITS*. Vol. 6 (1), 122-128.
- Sabat, T., Meilanani, S. dan Purnomo, W. (2017). Spatial Modeling of Infant Mortality Rate In South Central Timor Regency Using GWLR Method With Adaptive Bisquare Kernel and Gaussian Kernel. *Health Nations*. Vol. 1 (2), 90-98.
- Suyitno, Purhadi, Sutikno, dan Irhamah. (2016). Parameter Estimation of Geographically Weighted Trivariate Weibull Regression Model. *Applied Mathematical Science*. Vol. 10(18), 861-878.
- Yrigoyen, C. C., Rodriguez, I. G. dan Otero, J. V. 2007. Modeling Spatial Variations In Household Disposable Income With Geographically Weighted Regression. *MPRA Paper*. No. 1682.

ANALISIS PERBANDINGAN PERFORMA METODE ENSEMBLE DALAM MENANGANI *IMBALANCED MULTI-CLASS CLASSIFICATION*

Qorry Meidianingsih¹, Devi Eka Wardani Meganingtyas²

¹Program Studi Pendidikan Matematika, Fakultas MIPA, Universitas Negeri Jakarta

²Program Studi Matematika, Fakultas MIPA, Universitas Negeri Jakarta

e-mail: ¹qorrymeidianingsih@unj.ac.id, ²deviekawm@unj.ac.id

Received: December 5, 2021; Revised: February 23, 2022; Accepted: May 20, 2022; Published: June 30, 2022

Abstrak

Penelitian ini fokus pada membandingkan kinerja beberapa metode *ensemble* dalam mengatasi *imbalanced multi-class classification* dimana metode dekomposisi *one-versus-one* (OVO) diterapkan sehingga metode klasifikasi *Support Vector Machine* yang standar dapat digunakan. Data penelitian merupakan hasil bangkitan *software R* yang dirancang berdasarkan level persentase kelas minoritas, yaitu menjadi kategori *extreme*, *moderate*, dan *mild*. Selain itu, dirancang pula kombinasi jumlah kelas mayoritas dan minoritas yang mungkin terjadi sehingga terdapat sembilan jenis data simulasi. Hasil penelitian menunjukkan bahwa secara umum rata-rata ketepatan klasifikasi paling rendah diperoleh ketika data simulasi bersifat *moderate* dengan sebaran gugus data terbagi menjadi 2 kelas mayoritas dan 1 kelas minoritas. Metode *safe-level SMOTEBagging* memberikan performa yang paling baik bagi semua jenis data simulasi, terutama ketika gugus data bersifat ekstrim, yaitu ditunjukkan dengan nilai rata-rata akurasi yang diperoleh sebesar 98.60 persen. Kinerja metode klasifikasi SVM cukup baik dimana nilai rata-rata tingkat akurasi setiap kelas berkisar antara 67.80-98.60 persen.

Kata kunci: *imbalanced multi-class, ensemble method, bagging, underbagging, SMOTEBagging, safe-level SMOTE, support vector machine*

Abstract

This study focused on comparing the performance of several ensemble methods in handling imbalanced multi-class classification where the one-versus-one (OVO) decomposition method is applied so that the standard Support Vector Machine classification method can be used. The data were generated by R software which is designed based on the percentage level of the minority class, namely into "extreme", "moderate", and "mild" categories. In addition, a combination of the number of majority and minority classes that may occur is designed so that there are nine types of simulation data. The results showed that in general the lowest average classification accuracy was obtained when the simulation data was moderate with the distribution of the data set divided into 2 majority classes and 1 minority class. The safe-level SMOTEBagging method provides the best performance for all types of simulation data, especially when the data set is extreme, which is indicated by the average accuracy obtained of 98.60 percent. The performance of the SVM classification method is quite good where the average value of the accuracy level of each class ranges from 67.80-98.60 percent.

Keywords: *imbalanced multi-class, ensemble method, bagging, underbagging, SMOTEBagging, safe-level SMOTE, support vector machine*

PENDAHULUAN

Penelitian yang melibatkan data besar (*big data*) menjadi kajian yang banyak dipelajari oleh komunitas *data mining* dan *machine learning*. Skenario masalah yang ditemukan pun lebih beragam sehingga dapat memberikan berbagai pilihan topik, dimana salah satu yang mendapat perhatian yaitu mengenai kelas tidak seimbang. Berbagai penelitian telah dilakukan untuk menangani permasalahan klasifikasi ketika kelas tidak seimbang, namun banyak literatur yang lebih fokus pada kelas bersifat biner (memiliki dua kelas).

Permasalahan data dengan ketidakseimbangan pada *multi-class* mulai mendapatkan perhatian dari komunitas peneliti dalam beberapa tahun terakhir (Lango, 2019). Literatur mengenai *pattern recognition* telah banyak membahas teknik klasifikasi pola pada *multi-class*, namun sebagian besar teknik tersebut menerapkannya pada kondisi data yang seimbang, seperti klasifikasi dokumen tertulis dan pengenalan suara, meskipun dalam banyak penerapan di bidang lain data yang ditemukan lebih sering memiliki kelas yang tidak seimbang (Ghanem *et. al.*, 2010). Dalam menangani permasalahan ketidakseimbangan *multi-class*, kita akan dihadapkan pada gugus data dengan konfigurasi yang berbeda, seperti kemungkinan untuk memiliki tidak hanya satu kelas mayoritas tetapi beberapa dan sebaliknya dimana satu kelas mayoritas dan beberapa kelas minoritas (Esteves, 2020). Beberapa metode standar pun tidak dapat diterapkan secara langsung pada kasus ketidakseimbangan *multi-class*. Hal ini lah yang membuat permasalahan klasifikasi pada kasus *imbalanced multi-class* menjadi lebih rumit. Metode yang digunakan harus mencoba untuk menangkap dan menggali lebih jauh hubungan antara beberapa kelas yang tidak seimbang serta siap untuk bekerja dengan ketidakseimbangan yang lebih ekstrim karena mereka lebih mungkin terjadi dalam gugus data *multi-class* (Lango, 2019).

Dalam mengatasi permasalahan klasifikasi *multi-class*, metode dekomposisi merupakan salah satu yang populer, sedangkan metode *ensemble* dapat digunakan untuk menangani permasalahan ketidakseimbangan kelasnya. Metode dekomposisi diterapkan untuk membagi gugus data *multi-class* menjadi multipel sub-kelas biner sehingga nantinya akan menghasilkan pengklasifikasi biner sebanyak sub-kelas. Metode *ensemble* yang digunakan untuk menyeimbangkan jumlah amatan kelas pun beragam, di antaranya *boosting* dan *bagging*. Penyeimbangan jumlah amatan antara kelas mayoritas dan minoritas memungkinkan metode klasifikasi standar dapat diterapkan, salah satunya yaitu *Support Vector Machine*. Penelitian ini fokus kepada membandingkan kinerja beberapa metode *ensemble*, yaitu *underbagging*, *SMOTEBagging*, dan *safe-level SMOTEBagging*, dalam mengatasi *imbalanced multi-class classification* menggunakan metode dekomposisi *one-versus-one* (OVO). Dengan demikian hasil penelitian ini diharapkan dapat memberikan alternatif lain dalam mengatasi kasus klasifikasi untuk gugus data *multi-class* yang tidak seimbang.

METODOLOGI

Tinjauan Referensi

Pendekatan yang diperkenalkan untuk mengatasi permasalahan data dengan kelas tidak seimbang hampir sebagian besar hanya dirancang untuk skenario kelas biner. Beberapa metode tidak dapat diterapkan secara langsung pada kasus ketidakseimbangan *multi-class*. Solusi yang ditawarkan dalam mengatasi ketidakseimbangan *multi-class* cukup beragam, namun sebagian besar literatur mengarahkan pada metode dekomposisi kelas atau mengkonversi masalah *multi-class* menjadi sekumpulan sub-kelas biner. (Wang dan Yao, 2012). Metode dekomposisi tersebut pada awalnya digunakan pada penerapan klasifikasi

multi-class yang seimbang, namun Esteves (2020) memberikan penguatan bahwa metode tersebut dapat dipelajari oleh pengklasifikasi biner sekalipun kondisi datanya tidak seimbang.

One-versus-all (OVA) dan *One-versus-one* (OVO) merupakan dua pendekatan yang membagi masalah untuk memungkinkan hasil yang diperoleh menjadi lebih baik. Kedua metode tersebut sangat dikenal ketika melakukan dekomposisi kelas. OVA pertama kali didiskusikan sebagai sebuah teknik oleh Clark dan Boswell pada tahun 1991 (dalam Esteves, 2020). Ide dari metode OVA yaitu memberikan label pada salah satu kelas sebagai positif dan kelas lainnya sebagai negatif lalu menerapkan metode klasifikasi yang sama pada semua kelas dan membuat model klasifikasi berdasarkan data latih. Hasil dari metode tersebut adalah model klasifikasi multipel biner yang dapat dibandingkan hingga diperoleh yang terbaik. Metode *one-versus-one* (OVO) diperkenalkan oleh Knerr *et al.* pada tahun 1990 (dalam Azis *et al.*, 2017). Pada metode OVO, metode klasifikasi biner dibangun oleh setiap pasangan kelas. Pendekatan sederhana untuk menggabungkan semua hasil dari metode klasifikasi biner pada OVO yaitu menggunakan *majority voting* dimana contoh uji akan diklasifikasikan ke dalam kelas yang memiliki *vote* tertinggi (Ghanem *et. al.*, 2010).

Secara umum metode OVO lebih disukai untuk data dengan kelas yang tidak seimbang karena masalah biner yang dibangun oleh OVA bahkan lebih sangat tidak seimbang daripada masalah yang sebenarnya. Di sisi lain, metode OVO akan memakan banyak waktu dan biaya ketika jumlah kelas yang dimiliki semakin banyak karena akan menghasilkan pengklasifikasi sub-kelas yang sangat banyak dan memungkinkan untuk memberikan masalah yang rumit dalam pengerjaannya (Lango, 2019). Meskipun demikian, penelitian yang dilakukan oleh Rout dan Mishra (2018) memberikan hasil yang positif mengenai aplikasi metode OVO.

Penelitian tersebut berupa eksperimen untuk menganalisis permasalahan klasifikasi pada imbalanced *multi-class*. Metode OVO diterapkan pada gugus data penelitian sehingga dihasilkan sekumpulan kelas biner yang tidak seimbang. Beberapa metode ensemble berbasis boosting selanjutnya digunakan untuk menghasilkan gugus sub-kelas yang seimbang sehingga metode klasifikasi biner dapat diterapkan. Penggunaan metode OVO dipilih pada penelitian tersebut dengan merujuk pada penelitian yang dilakukan oleh Saez *et al.* (2014) yang memberikan hasil bahwa metode OVO menunjukkan performa yang lebih baik dan pengklasifikasi yang lebih robust ketika berhadapan dengan data yang memiliki noise.

Dalam menangani permasalahan ketidakseimbangan kelas, pendekatan solusi yang telah banyak dikenal yaitu pada level data preprocessing dan modifikasi algoritma metode klasifikasi standar. Data preprocessing merupakan tahap mengubah sebaran data dari semua kelas menjadi seimbang mungkin. Gugus data yang telah melewati tahap ini selanjutnya dapat diterapkan suatu pengklasifikasi tunggal dan diukur kinerjanya. Salah satu metode yang menarik perhatian adalah metode ensemble, yaitu menggabungkan banyak pengklasifikasi (classifier) tunggal yang kemudian hasil prediksi dari masing-masing pengklasifikasi digabungkan dengan proses voting. Salah satu algoritma metode ensemble yang dapat digunakan dalam skenario *multi-class* adalah bagging. Apabila dibandingkan dengan pengklasifikasi tunggal, hasil prediksi bagging hampir selalu lebih akurat (Zhou, 2012).

Tanha *et al.* (2020) mengevaluasi berbagai metode berbasis algoritma boosting pada dua tipe gugus data *multi-class* yang tidak seimbang, yaitu konvensional dan data besar. Metode OVO diterapkan untuk membagi gugus data *multi-class* menjadi multipel sub-kelas biner. Hasil penelitian menunjukkan bahwa tidak ada algoritma boosting yang

menunjukkan performa baik pada gugus data besar yang memiliki tingkat ketidakseimbangan kelasnya tinggi. Nilai G-mean yang sangat kecil menunjukkan bahwa metode boosting tidak dapat mempelajari beberapa kelas pada gugus data. Oleh karena itu, algoritma boosting tidak cocok digunakan pada data besar dengan rasio ketidakseimbangan kelas yang lebih tinggi. Meskipun begitu, pada penelitian tersebut ditemukan bahwa teknik oversampling cukup efektif untuk gugus data multi-class yang tidak seimbang.

Kajian beberapa teori dan hasil penelitian sebelumnya yang terkait dengan penelitian ini memberikan gambaran bahwa algoritma *bagging* sebagai salah satu bagian dari metode *ensemble* belum banyak dilibatkan dalam topik penelitian *imbalanced multi-class classification*. Penggunaan metode dekomposisi *one-versus-one* dan penerapan *support vector machine* sebagai metode klasifikasi diharapkan mampu memberikan hasil prediksi yang lebih baik. Hal ini didukung dengan alasan bahwa *Support Vector Machine* merupakan salah satu metode yang kini banyak mendapat perhatian sebagai *state of the art* dalam klasifikasi *pattern recognition* (Azis *et al.*, 2017).

UnderBagging

Barandela *et al.* (2003) memperkenalkan metode *UnderBagging* pada tahun 2003. Metode ini merupakan gabungan antara teknik pengambilan contoh *undersampling* dan teknik *ensemble bagging*. Ide dari metode ini yaitu membuat beberapa data latih sebanyak k kali dimana seluruh amatannya berasal dari kelas minoritas dan menarik secara acak amatan kelas mayoritas dengan jumlah yang sama atau dengan teknik tanpa pengembalian (Blaszczynski *et al.*, 2013). Selanjutnya metode klasifikasi diterapkan pada setiap data latih tersebut sehingga diperoleh k buah klasifikasi sehingga hasil prediksi gabungannya ditentukan berdasarkan *majority vote*.

SMOTEBagging

Synthetic Minority Oversampling Technique (SMOTE) merupakan salah satu teknik *oversampling* yang diperkenalkan oleh Chawla *et al.* pada tahun 2002. Oleh karena itu, *SMOTEBagging* merupakan kombinasi antara metode SMOTE dengan algoritma *bagging* yang diperkenalkan oleh Breiman (1996). Ide dari *SMOTEBagging* adalah menerapkan SMOTE sebelum pemodelan pada setiap gugus data yang diperoleh melalui proses *bootstrap* agar kelas menjadi lebih seimbang. Tujuan dari metode ini adalah menciptakan model yang *powerful* dalam mengklasifikasi data yang tidak seimbang tanpa mengorbankan akurasi secara keseluruhan (Hanifah, 2015).

Prosedur SMOTE menurut Chawla *et al.* (2002) adalah sebagai berikut:

1. Tentukan k tetangga terdekat untuk setiap amatan kelas minoritas.
2. Pilih 1 dari k tetangga terdekat secara acak.
3. Tentukan nilai selisih antara amatan kelas minoritas dan tetangga terdekat yang terpilih.
4. Kalikan hasil pada tahap 3 dengan bilangan acak antara 0 dan 1.
5. Tambahkan hasil di poin 4 dengan amatan kelas minoritas yang awal.
6. Hasil tahap 5 merupakan data sintesis yang dibangkitkan.

Safe-level SMOTEBagging

Metode ini sebenarnya mengadaptasi algoritma yang diterapkan pada *SMOTEBagging*. Perbedaan antara metode ini dengan *SMOTEBagging* terletak pada cara membangkitkan data sintesisnya dimana dilakukan berdasarkan pada kriteria wilayah aman yang diperkenalkan oleh Bunkhumpornpat *et al.* (2009).

Ukuran Kinerja Metode Klasifikasi

Dalam mengevaluasi kinerja metode klasifikasi pada kasus *multi-class*,

prosedurnya dapat mengikuti pada kasus kelas biner. Lapalme dan Sokolova (2009) menjabarkan lebih rinci penerapan perhitungan akurasi berdasarkan *confusion matrix* ketika klasifikasi diterapkan pada *multi-class*.

$$\text{Average Accuracy} = \frac{\sum_{i=1}^l \frac{TP_i + TN_i}{TP_i + FP_i + FN_i + TN_i}}{l}$$

Keterangan:

TP_i : *True positive* atau banyaknya amatan kelas positif yang diklasifikasikan dengan tepat pada kelas ke- i

TN_i : *True negative* atau banyaknya amatan kelas negatif yang diklasifikasikan dengan tepat pada kelas ke- i

FP_i : *False positive* atau banyaknya amatan kelas negatif dan diklasifikasikan ke dalam kelas positif pada kelas ke- i

FN_i : *False negative* atau banyaknya amatan kelas positif dan

diklasifikasikan ke dalam kelas negatif pada kelas ke- i

l : jumlah kelas

Metode Analisis

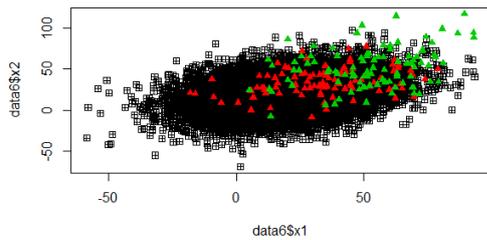
Penelitian ini menggunakan data simulasi, yaitu data hasil *generate* melalui *software R*. Variabel penelitian terdiri dari variabel prediktor X_1 dan X_2 yang bersifat numerik, sedangkan peubah responnya terdiri dari tiga kelas, yaitu 1, 2, dan 3. Terdapat beberapa skema data simulasi yang ditentukan berdasarkan tingkat persentase kelas minoritas serta banyaknya kelas minoritas pada satu gugus data simulasi. Hal tersebut merupakan implikasi dari kondisi *multi-class* yang memungkinkan dalam satu gugus data terdapat lebih dari satu kelas minoritas atau kelas mayoritas seperti yang disajikan pada Tabel 1. Dengan demikian, data simulasi yang dibangkitkan diharapkan dapat menangkap kemungkinan-kemungkinan tersebut.

Tabel 1. Skema Data Simulasi

Jenis Data	Persentase (%)			Keterangan
	Kelas 1	Kelas 2	Kelas 3	
<i>Extreme</i>	75	24	1	1 mayoritas, 1 minoritas
	50	49	1	2 mayoritas, 1 minoritas
	98	1	1	1 mayoritas, 2 minoritas
<i>Moderate</i>	75	20	5	1 mayoritas, 1 minoritas
	48	47	5	2 mayoritas, 1 minoritas
	90	5	5	1 mayoritas, 2 minoritas
<i>Mild</i>	65	25	10	1 mayoritas, 1 minoritas
	45	45	10	2 mayoritas, 1 minoritas
	80	10	10	1 mayoritas, 2 minoritas

Secara umum, terdapat 3 jenis data simulasi, yaitu *extreme*, *moderate*, dan *mild*. Kelas 1 ditetapkan sebagai kelas mayoritas, sedangkan kelas 3 ditentukan sebagai kelas minoritas. Kelas 2 ditentukan sebagai kelas minoritas ke-2 ketika dalam satu gugus data terdapat dua kelas minoritas, atau kelas mayoritas ke-2 jika hanya terdapat 1 kelas minoritas dalam gugus data. Bentuk sebaran ketiga kelas

dirancang agar saling beririsan. Sebaran tersebut dipilih untuk melihat performa metode penyeimbang kelas, seperti SMOTE dan *safe-level* SMOTE. Dengan demikian, sebaran data yang beririsan akan menentukan apakah data sintesis akan memberikan hasil klasifikasi yang baik atau sebaliknya.



Gambar 1. Contoh *Scatter Plot* Data Simulasi *Extreme* (1 mayoritas, 2 minoritas)

Tahapan analisis secara umum adalah sebagai berikut.

1. Membangkitkan data simulasi menggunakan *software R* dengan tahapan sebagai berikut.
 - 1) Membangkitkan 10.000 amatan populasi menggunakan bilangan acak normal multivariat.
 - 2) Membagi seluruh amatan populasi untuk kelas 1, kelas 2, dan kelas 3 sesuai dengan proporsi yang telah ditentukan pada Tabel 1.
 - 3) Tahap 1 dan 2 dilakukan untuk jenis data *extreme*, *moderate*, dan *mild* sehingga diperoleh 9 gugus data simulasi.
2. Membagi setiap gugus data simulasi menjadi data latih (80%) dan data uji (20%). Pemilihan data latih dan uji dilakukan berdasarkan teknik pengambilan sampel acak sederhana. Total terdapat sembilan buah gugus data simulasi.
3. Membentuk sub-kelas atau subset biner pada setiap data latih.
4. Menerapkan metode *Undersampling*, *SMOTE*, dan *safe-level SMOTE* pada masing-masing sub-kelas untuk membuat jumlah amatan di semua kelas menjadi seimbang.
5. Menerapkan metode klasifikasi *Support Vector Machine* dengan fungsi kernel radial basis pada setiap sub-kelas.
6. Menentukan hasil prediksi berdasarkan model SVM yang telah terbentuk menggunakan data uji.

7. Ulangi tahap 4-6 sebanyak 50 ($k=50$) kali dengan pemilihan sampel berulang menggunakan teknik *bootstrap*.
8. Lakukan *majority voting* berdasarkan hasil yang diperoleh pada tahap 7. Tahap ini disebut juga sebagai *bagging* atau *bootstrap aggregating*.
9. Hasil prediksi berdasarkan *bagging* pada setiap sub-kelas selanjutnya digabung dan dicari kembali nilai *majority voting*-nya. Hasil tersebutlah yang dijadikan sebagai hasil prediksi akhir.

HASIL DAN PEMBAHASAN

Penerapan Metode Dekomposisi OVO

Metode dekomposisi *one-versus-one* (OVO) diterapkan pada data latih yang telah terpilih secara acak. Andaikan kelas biner yang baru dinotasikan dengan kelas 1 dan 0, maka proses dekomposisi 3 kelas awal (kelas 1, 2, 3) akan menghasilkan 3 kelas biner seperti yang disajikan pada Tabel 2 berikut ini.

Tabel 2. Dekomposisi Kelas

Kelas Biner (Subset) ke-	Kelas 1	Kelas 0
1 (Subset A)	1	2
2 (Subset B)	1	3
3 (Subset C)	2	3

Gambaran penentuan hasil prediksi berdasarkan proses *bagging* dan dekomposisi dapat dilihat pada tabel 3 dan 4. Keseluruhan proses tersebut diterapkan pada seluruh jenis data simulasi.

Tabel 3 Hasil Prediksi Kelas berdasarkan Metode *Ensemble*

Amatan Data Uji ke-	<i>Undersampling</i>				SMOTE				<i>Safe-level SMOTE</i>			
	<i>Bootstrap ke-</i>											
	1	2	...	50	1	2	...	50	1	2	...	50
1												
2												
3												
⋮												
2000												

Tabel 4 Hasil Prediksi Kelas berdasarkan Metode Dekomposisi

Amatan Data Uji ke-	<i>UnderBagging</i>			Hasil Prediksi Kelas (<i>UnderBagging</i>)	SMOTE <i>Ba</i> gging			Hasil Prediksi Kelas (SMOTE <i>Ba</i> gging)	<i>Safe-level SMOTEBa</i> gging			Hasil Prediksi Kelas (<i>Safe-level SMOTEBa</i> gging)
	1 vs 2	1 vs 3	2 vs 3		1 vs 2	1 vs 3	2 vs 3		1 vs 2	1 vs 3	2 vs 3	
1												
2												
3												
⋮												
2000												

Performa Metode Klasifikasi

Performa metode klasifikasi SVM dapat dilihat berdasarkan hasil prediksi model berdasarkan data uji. Tabel 5, 6, dan 7 menunjukkan persentase rata-rata ketepatan setiap kelas dalam melakukan prediksi atau klasifikasi. Notasi UB ditujukan untuk *underbagging*, SMB untuk *SMOTEBagging*, dan SFB untuk *safe-level SMOTEBagging*.

Tabel 5. Persentase (%) rata-rata ketepatan klasifikasi setiap kelas berdasarkan tipe data *extreme*

Skema Data Simulasi	UB	SMB	SFB
1 kelas mayoritas, 1 kelas minoritas	73.30	75.77	83.33
2 kelas mayoritas, 1 kelas minoritas	68.73	71.70	75.33
1 kelas mayoritas, 2 kelas minoritas	76.97	81.23	98.60

Tabel 6. Persentase (%) rata-rata ketepatan klasifikasi setiap kelas berdasarkan tipe data *moderate*

Skema Data Simulasi	UB	SMB	SFB
---------------------	----	-----	-----

1 kelas mayoritas, 1 kelas minoritas	73.63	76.10	83.20
2 kelas mayoritas, 1 kelas minoritas	67.97	67.80	74.37
1 kelas mayoritas, 2 kelas minoritas	78.83	79.03	93.67

Tabel 7. Persentase (%) rata-rata ketepatan klasifikasi setiap kelas berdasarkan tipe data *mild*

Skema Data Simulasi	UB	SMB	SFB
1 kelas mayoritas, 1 kelas minoritas	73.80	77.13	77.50
2 kelas mayoritas, 1 kelas minoritas	70.87	71.53	74.87
1 kelas mayoritas, 2 kelas minoritas	76.30	76.60	88.37

Secara umum, rata-rata ketepatan klasifikasi paling rendah diperoleh ketika data simulasinya bersifat *moderate* dengan sebaran gugus data terbagi menjadi 2 kelas mayoritas dan 1 kelas minoritas (tipe ke-2). Sekalipun persentase kelas minoritas telah ditingkatkan (bersifat *moderate*), namun penerapan metode *Undersampling*, SMOTE, dan *safe-level SMOTE* dalam

menangani permasalahan kelas tidak seimbang terlihat kurang berhasil dalam memberikan model klasifikasi yang baik pada tipe data tersebut. Hal tersebut berbeda ketika tipe data yang digunakan bersifat *extreme*. Metode *SMOTEBagging* dan *safe-level SMOTEBagging* mampu menghasilkan rata-rata tingkat akurasi yang tinggi dimana nilainya secara berturut-turut sebesar 81.23 persen dan 98.60 persen. Berdasarkan hasil tersebut, terlihat bahwa metode *safe-level SMOTEBagging* memberikan performa yang sangat baik ketika gugus data bersifat ekstrim. Kinerja yang konsisten juga bisa dilihat pada jenis data simulasi lainnya. Rata-rata tingkat akurasi setiap kelas ketika metode *safe-level SMOTEBagging* diterapkan menunjukkan nilai yang cukup tinggi. Ide untuk membangkitkan data sintesis pada wilayah yang *safe* masih memberikan hasil yang positif dibandingkan dengan metode *SMOTEBagging* dan *UnderBagging*.

KESIMPULAN

Berdasarkan hasil penelitian, metode *safe-level SMOTEBagging* memberikan performa yang paling baik bagi semua jenis data simulasi, terutama ketika gugus data bersifat ekstrim, yaitu ditunjukkan dengan nilai rata-rata akurasi yang diperoleh sebesar 98.60 persen. Performa metode klasifikasi SVM dapat dilihat dari nilai rata-rata tingkat akurasi setiap kelas yang masih cukup tinggi (berkisar antara 67.80-98.60 persen). Dengan kata lain, strategi dekomposisi *one-versuse-one* cukup berhasil memecahkan masalah *multi-class* sehingga SVM standar dapat langsung diterapkan.

DAFTAR PUSTAKA

Azis, A. I. S., Suhartono, V., & Himawan, H. (2017). Model Multi-class SVM Menggunakan Strategi 1V1 untuk Klasifikasi Wall-Following Robot Navigation Data. *Jurnal Teknologi Informasi*, 13(2).

Barandela J., Sanchez, J.S., Garcia, V., & Rangel, E. (2003). Strategies for

Learning in Class Imbalance Problems. *The Journal of the Pattern Recognition Society*, 36, 849-851.

Blaszczynski, J., Stefanowski, J., & Idkowiak, L. (2013). Extending bagging for imbalanced data. *Proc. of the 8th CORES 2013, Springer Series on Advances in Intelligent Systems and Computing*, 226, 269-278.

Breiman, L. (1996). Bagging Predictors. *Machine Learning*, 24, 123-140. <https://doi.org/10.1023/A:1018054314350>.

Bunhumpornpat C., Sinapiromsaran K., Lursinsap C. (2009) Safe-Level-SMOTE: Safe-Level-Synthetic Minority Over-Sampling TEchnique for Handling the Class Imbalanced Problem. In: Theeramunkong T., Kijirikul B., Cercone N., Ho TB. (eds) *Advances in Knowledge Discovery and Data Mining. PAKDD 2009. Lecture Notes in Computer Science*, 5476, 475-482. https://doi.org/10.1007/978-3-642-01307-2_43.

Chawla, N. V., Bowyer, K. W., Hall, L.O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357. <https://doi.org/10.1613/jair.953>.

Esteves, V. M. S. (2020). Techniques to deal with imbalanced data in multi-class problems: A review of existing methods. *Faculdade de Engenharia da Universidade do Porto*.

Ghanem, A. S., Venkatesh, S., & West, G. (2010). Multi-class pattern classification in imbalanced data. *Proceedings: 20th International Conference on Pattern Recognition, IEEE, Los Alamitos, Calif.*, 2881-2884. <https://doi.org/10.1109/ICPR.2010.706>.

Hanifah FS. 2015. Penerapan algoritme *SMOTEBagging* dalam penyusunan pohon keputusan dan regresi logistik untuk kajian kredit macet [tesis]. Bogor (ID): Institut Pertanian Bogor.

Lango, M. (2019). Tackling the Problem of Class Imbalance in Multi-class

- Sentiment Classification: An Experimental Study. *Scienco*, 44(2). <https://doi.org/10.2478/fcds-2019-0009>.
- Rout, N., Kuhoo, & Mishra, D. (2018). Analysing the Multi-class Imbalanced Datasets using Boosting Methods and Relevant Information. *International Journal of Pure and Applied Mathematics*, 120(6), 6171-6191.
- Saez, J. A., Galar M., Luengo, J., & Herrera, F. (2014). Analysing the presence of noise in multi-class problems: alleviating its influence with the One-vs-One decomposition. *Knowledge and Information Systems*, 38(1), 179-206. <https://doi.org/10.1007/s10115-012-0570-1>.
- Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., & Asadpour, M. (2020). Boosting methods for multi-class imbalanced data classification: an experimental review. *Journal of Big Data*, 7(70). <https://doi.org/10.1186/s40537-020-00349-y>.
- Wang, S. & Yao, X. (2012). Multi-class Imbalance Problems: Analysis and Potential Solutions. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 42(4). <https://doi.org/10.1109/TSMCB.2012.2187280>.
- Zhou, Z. (2012). *Ensemble Methods: Foundations and Algorithms*. CRC Press

KLASIFIKASI KARAKTERISTIK KEMISKINAN DI PROVINSI BENGKULU TAHUN 2020 MENGGUNAKAN METODE POHON KLASIFIKASI GABUNGAN

Winalia Agwil¹, Dian Agustina², Herlin Fransiska³, Nurul Hidayati⁴

^{1,2,3,4} Program Studi Statistika, FMIPA Universitas Bengkulu, Bengkulu

e-mail: ¹winaliaagwil@unib.ac.id

Received: December 31, 2021; Revised: April 5, 2022; Accepted: June 2, 2022; Published: June 30, 2022

Abstrak

Kemiskinan adalah masalah yang mendesak diatasi baik pada tingkat nasional maupun global, yang diindikasikan sebagai salah satu prioritas utama dalam agenda dunia tentang Tujuan-tujuan Pembangunan Berkelanjutan atau *Sustainable Development Goals* (SDGs). Penanganan kemiskinan yang efektif akan membantu penyelesaian permasalahan dunia yang lain seperti permasalahan kelaparan, kesehatan, kesejahteraan, pendidikan, air bersih dan sanitasi. Studi ini bertujuan untuk mengklasifikasikan karakteristik kemiskinan rumah tangga di Propinsi Bengkulu, berdasarkan data hasil Survei Sosial Ekonomi Nasional (SUSENAS) 2020. Metode analisis data untuk mengidentifikasi karakteristik rumah tangga miskin dalam studi ini menggunakan *Classification and Regression Tree* (CART) – dengan menerapkan model random forest untuk mengatasi ketidakseimbangan data. Berdasarkan pemodelan tersebut, studi ini menemukan bahwa terdapat tiga variabel utama yang mencirikan rumah tangga miskin di propinsi Bengkulu, yaitu jumlah anggota rumah tangga, ijazah terakhir kepala rumah tangga dan luas lantai rumah. Temuan ini dapat digunakan sebagai dasar untuk identifikasi rumah tangga miskin, sehingga program-program bantuan diharapkan lebih tepat sasaran di masa mendatang.

Kata Kunci: Kemiskinan, Unbalanced dataset, CART, Random forest, Xgboost

Abstract

Poverty is a pressing issue at both country and global level, indicated as one of main priorities in the global agenda on Sustainable Development Goals (SDGs). Tackling poverty effectively will help solve other critical issues such as hunger, health, welfare, education, clean water, and sanitation. The aim of this study is to classify the main characteristics of poor households in Province of Bengkulu, using the data from the results of National Socioeconomic Survey (SUSENAS) 2020. Method of data analysis in identifying the poor household characteristics in the present study uses Classification and Regression Tree (CART) – specifically by applying a forest random model in attempt to adjust the unbalanced datasets. Main findings from this modelling application suggest that there are three variables that mainly characterized poor households in Bengkulu, namely the number of household members, the last education certificate of the household head, and the size of the house floor area. This finding is useful for a basis in identifying the poor households, thus intervention programs designed to assist the poor is expected to be more well-targeted in the future.

Keywords: Poverty, Unbalance datasets, CART, Random forest

PENDAHULUAN

Kemiskinan merupakan isu mendesak untuk diatasi di tingkat nasional dan deklarasi *Millenium Development Goals* (MDGs) yang salah satu targetnya adalah mengurangi penduduk miskin dunia. Era MDGs berhasil mencapai tujuannya yakni mengatasi kemiskinan hingga 50%. Selanjutnya memasuki era *Sustainable Development Goals* (SDGs) yang dicetus untuk meneruskan dan memantapkan capaian MDGs agar berlanjut. SDGs memprioritaskan “*no poverty*”, karena mengatasi kemiskinan secara efektif akan sejalan dengan upaya-upaya mengatasi permasalahan dunia lainnya seperti: dunia tanpa kelaparan, kesehatan yang baik, kesejahteraan, pendidikan berkualitas, air bersih dan sanitasi (Ishartono dan Raharjo, 2016). Pada tahun 2020, Provinsi Bengkulu merupakan provinsi dengan persentase penduduk miskin kedua tertinggi di pulau Sumatera setelah Aceh, yaitu mencapai 14,43%. Mengetahui karakteristik rumah tangga miskin di Provinsi Bengkulu menjadi penting untuk dikaji, sebagai acuan dalam penyaluran program pengetasan kemiskinan sehingga program pemerintah dapat dilakukan secara tepat sasaran dan efisien.

Pengkajian tentang kemiskinan dapat dilihat dari unit paling kecil yakni rumah tangga. Menurut BPS, rumah tangga dikategorikan sebagai miskin, jika nilai rata-rata pengeluaran konsumsi per kapita per bulan dibawah garis kemiskinan. Dan sebaliknya, jika nilai rata-rata pengeluaran konsumsi per kapita per bulan berada diatas garis kemiskinan maka rumah tangga tersebut dapat di kategorikan tidak miskin. Penentuan karakteristik dari rumah tangga miskin dan tidak miskin dapat dilakukan dengan analisis klasifikasi. Metode klasifikasi yang paling umum digunakan antara lain, regresi logistik, klasifikasi pohon, K-tetangga terdekat, jaringan saraf tiruan, dan Support Vector Machine (SVM). Setiap metode memiliki syarat/batasan untuk dapat digunakan juga memiliki kelebihan dan kekurangan. Secara umum, permasalahan ketidakseimbangan kelas data menyebabkan analisis klasifikasi pohon tunggal akan menghasilkan pohon yang

kurang stabil dimana ketika data training mengalami perubahan kecil dapat memberikan perubahan yang signifikan pada pohon yang dihasilkan (Sutton, 2005).

Ispriyanti, Prahutama dan Mustafid (2019) juga eneliti pengklasifikasian kemiskinan di Kota Semarang dengan menerapkan metode algoritma QUEST. Metode QUEST termasuk salah satu dari banyak metode pohon klasifikasi tunggal. Penelitian tersebut menunjukkan perbandingan antara jumlah rumah tangga miskin dan tidak miskin sebesar 47:883. Hal ini mengindikasikan adanya ketidakseimbangan kelas data yang potensial mengakibatkan kesalahan klasifikasi saat memprediksi kelas data. Hal ini misalnya hasil klasifikasi pada awalnya diperoleh klasifikasi rumah tangga miskin, namun setelah di prediksi klasifikasi berubah menjadi rumah tangga tidak miskin.

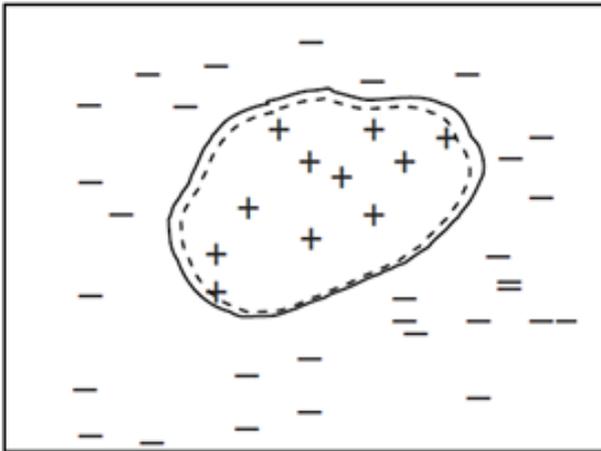
Kondisi ketidakseimbangan data perlu untuk ditangani sehingga memperbaiki performa model klasifikasi, penulis tertarik untuk menangani permasalahan ini dengan menerapkan metode Synthetic Minority Over-sampling Technique (SMOTE) yang merupakan solusi pada level data dan juga menerapkan metode pohon klasifikasi gabungan seperti Random Forest (RF) dan Extreme Gradient Boosting (Xgboost) untuk pemodelan klasifikasi.

Tujuan penelitian ini adalah untuk mengidentifikasi karakteristik rumah tangga miskin dan tidak miskin di Provinsi Bengkulu, kemudian melakukan pengklasifikasian status kemiskinan rumah tangga dengan menggunakan metode CART, Random Forest dan Extreme Gradient Boosting (Xgboost). Selain itu, masalah ketidakseimbangan data ditanggulangi dengan metode SMOTE yang kemudian datanya akan digunakan untuk pengklasifikasian rumah tangga dengan menggunakan metode CART, Random Forest Extreme Gradient Boosting (Xgboost). Dimana output dari penelitian adalah menghasilkan model pohon klasifikasi yang menggambarkan karakteristik kemiskinan di Provinsi Bengkulu. Model dibentuk dengan menggunakan data hasil SUSENAS Maret 2020 modul KOR dan Modul Konsumsi.

METODOLOGI

Ketidakseimbangan Data

Ketidakseimbangan data terjadi apabila pada suatu dataset terdapat suatu kategori (kelas) data yang mendominasi kategori yang lainnya. Kategoridata yang memiliki jumlah observasi yang lebih banyak disebut dengan kelas mayoritas (kelas negatif), sedangkan kategori data yang memiliki jumlah observasi lebih sedikit disebut dengan kelas minoritas (kelas positif) (Chawla et al. 2002). Ketidakseimbangan data diilustrasikan pada Gambar 1.



Gambar 1 Data Tidak Seimbang

Sumber: Chawla et al. 2002

Pada Gambar 1, dapat dilihat tanda “+” menggambarkan obsevasi yang berasal dari kategori minoritas dan “-“ mewakili observasi dari kelas mayoritas.

Penanganan ketidakseimbangan data dapat dilakukan dengan menggunakan Synthetic Minority Oversampling Technique (SMOTE). Metode SMOTE pertama kali dikenalkan oleh Chawla et.al (2002) yang merupakan metode penanggulangan ketidakseimbangan data dengan melakukan resampling pada data.

Metode SMOTE adalah proses penambahan data baru pada kelas minoritas agar jumlah observasinya sebanding dengan kelas mayoritas, yang dilakukan berdasarkan informasi tetangga terdekat (k-nearest neighbor). Terdapat dua tahapan dalam metode ini, yang pertama tahapan menentukan tetangga terdekat untuk setiap observasi yang berada pada kelas minoritas dengan menggunakan jarak Euclidean jika data berupa numerik dan Value Difference Metric (VDM) jika data kategorik. Tahapan

kedua adalah pembuatan data sintetik. Pada data numerik, data baru dihitung dengan persamaan $x^* = x_i + (bilangan\ acak(0 - 1) \cdot selisih(x_i, x_{ij}))$. Sedangkan untuk data kategorik, observasi baru merupakan mayoritas nilai dari k-tetangga terdekatnya (Chawla et al. 2002).

Classification and Regression Tree (CART)

Metode CART merupakan salah satu pohon keputusan populer yang dapat digunakan pada variabel respon numerik maupun kategorik. Jika variabel respon berupa data kategorik dinamakan pohon klasifikasi, sedangkan variabel respon yang berupa data numerik dinamakan pohon regresi. Pembentukan pohon klasifikasi maupun pohon regresi dilakukan dengan proses rekursif biner pada gugus data sehingga pada pemilahan terakhir diperoleh nilai variabel respon pada setiap simpul yang terbentuk lebih homogen (Breiman et al. 1984).

Lewis (2000) menyebutkan metode ini sebagai klasifikasi *binary recursive partitioning*, karena setiap simpul yang dihasilkan disekat atau dipisahkan menjadi dua simpul anak. Tahapan penyekatan tersebut dilakukan sampai terpenuhi kriteria pemberhentian yang ditetapkan. Algoritma CART secara umum digambarkan dalam tahapan berikut (Breiman et. Al. 1984):

1. Menemukan pemisah terbaik pada setiap variabel prediktor yang digunakan dalam model. Setiap variabel dengan nilai K yang berbeda memiliki k-1 dengan kemungkinan pemisah. Pemisah terbaik dipilih berdasarkan kriteria *splitting* yaitu *Gini's impurity*. Setiap variabel prediktor memiliki satu nilai pemisah terbaik. Berikut adalah formula indeks Gini:

$$i(t) = - \sum_{j=1}^J p(j|t) \log_2 p(j|t) \quad (1)$$

$p(j|t)$ merupakan peluang observasi kelas j pada *simpul* t. Evaluasi pemisah s pada simpul t dapat dilakukan dengan melihat nilai *Goodness of split*.

2. Menemukan variabel prediktor terbaik (satu dari semua variabel prediktor yang tersedia) yang dapat memisahkan dataset pada simpul sebelumnya sehingga lebih homogen pada sub-simpul. Pemilihan variabel prediktor terbaik juga dilakukan dengan menggunakan indeks Gini.
3. Lakukan penyekatan dengan menggunakan variabel prediktor pada tahapan (2), periksa apakah sudah memenuhi kriteria pemberhentian, jika tidak maka lakukan kembali tahapan (1).

Random Forest

Random Forest merupakan pengembangan metode *Bagging (Bootstrap Aggregating)*. Metode ini bertujuan memperbaiki performa klasifikasi tunggal dengan nilai akurasi yang rendah (Wezel dan Potharst, 2007). Pengembangan metode *bagging* tersebut terletak pada proses resampling. Resampling pada metode *random forest* tidak hanya pada pengacakan observasinya, tetapi juga pada variabel prediktor, sehinggasetiap proses resampling yang dilakukan memuat sampel yang berbeda. Akibatnya ukuran dan bentuk pohon klasifikasi juga berbeda (Liaw dan Weiner, 2002). Berikut adalah tahapan *Random Forest* secara umum (Sartono dan Syafitri, 2010):

1. a. tahapan resampling

Tahapan ini melakukan pembentukan sampel baru. Sampel dibentuk dari proses penarikan sampel acak dengan pemulihan sebanyak n data dari data training.

- b. tahapan random sub-setting

Tahapan ini membentuk model pohon klasifikasi (salah satunya CART) dengan data yang telah diperoleh pada tahapan 1.a, namun setiap proses pemisahan dilakukan pemilihan secara acak $m < d$ variabel penjelas. d adalah banyaknya variabel prediktor yang digunakan.

- c. ulangi langkah 1.a dan 1.b sebanyak B kali sehingga diperoleh B pohon klasifikasi (CART)

2. Majority Vote

Melakukan pendugaan gabungan berdasarkan hasil prediksi mayoritas dari B prediksi yang terbentuk.

Extreme Gradient Boosting (Xgboost)

Extreme Gradient Boosting (Xgboost) merupakan metode *ensemble* yang mirip dengan *gradient boosting*, namun lebih efisien karena terdapat parameter regulasi yang dapat mengontrol kompleksitas model (Chen dan Guestrin, 2016). Secara sederhana tahapan dalam pembentukan model *gradient boosting* adalah sebagai berikut ((Friedman, 2000):

- a. Membentuk model awal berupa nilai konstan yang memenuhi ketentuan berikut:

$$F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, \gamma) \quad (2)$$

- b. Untuk tahapan $m = 1$ sampai dengan M , maka lakukan: 1). Hitung residual atau error dari model sebelumnya untuk setiap observasi. 2). Pembentukan model pohon dengan menggunakan data x sebagai prediktor dan error pada tahapan sebelumnya sebagai respon. 3). Menghitung prediksi model. 4). Ulangi langkah 2.
- c. pembentukan model akhir

$$\hat{y}_i^{(M)} = F_0(x) + \sum_{m=1}^M \gamma_m F_m(x) \quad (3)$$

Algoritma *Xgboost* dapat digunakan dalam regresi maupun klasifikasi, dan dikenal memiliki performa yang baik dibanding metode *gradient boosting*.

Evaluasi Keباikan Model

Keباikan evaluasi dapat dianalisis dengan *Confusion Matrix* (Han, Kamber, dan Pei, 2012). Misalkan terdapat dua kelas data yaitu kelas positif dan kelas negatif, maka prediksi dapat diklasifikasikan sebagai berikut:

Tabel 1 Confusion Matrix

Prediksi	Aktual	
	Positif	Negatif
Positif	<i>True Positive</i> (TP)	<i>False Positive</i> (FP)
Negatif	<i>False Negative</i> (FN)	<i>True Negative</i> (TN)

Sumber: Han, Kamber, dan Pei, 2012

Kebaikan model klasifikasi dapat diukur dengan formula berikut:

1. $Accuracy = \frac{TP+TN}{(TP+TN+FN+FP)}$
2. $Sensitivity\ or\ recall = \frac{TP}{TP+FN}$
3. $Specificity = \frac{TN}{TN+FP}$
4. $Precision = \frac{TP}{TP+FP}$
5. $F - Score = \frac{2(recall)(precision)}{recall+precision}$.

Evaluasi kebaikan model klasifikasi dapat juga dilihat dari nilai AUC, yaitu nilai yang menggambarkan luas area di bawah kurva ROC dengan nilai antara 0 hingga 1. Sedangkan, Kurva ROC adalah kurva yang menggambarkan kebaikan model klasifikasi yang disajikan dalam dua dimensi. Kurva tersebut memuat nilai persentase *False Positive* (1- *Specificity*) dengan persentase *True Positive* (Fawcett 2006).

Metode Analisis

Data yang digunakan dalam penelitian ini ialah data hasil Survei Sosial Ekonomi Nasional (SUSENAS) Kor tahun 2020 dan SUSENAS Modul Konsumsi 2020. Objek pengamatan dalam penelitian ini adalah rumah tangga. Variabel respon pada penelitian ini adalah status kemiskinan rumah tangga yang dikategorikan menjadi rumah tangga miskin jika pengeluaran konsumsi per kapita per bulan berada di bawah garis kemiskinan dan jika pengeluaran konsumsi per kapita per bulan melebihi garis kemiskinan maka rumah tangga dikategorikan rumah tangga tidak miskin. Per Maret 2020, nilai garis kemiskinan Provinsi Bengkulu adalah Rp 527.031 per bulan per orang.

Variabel prediktor yang digunakan ialah: Jenis Kelamin Kepala Rumah Tangga (X1), Usia Kepala Rumah Tangga (X2), Banyaknya Anggota Rumah Tangga (X3), Ijazah Tertinggi Kepala Rumah Tangga (X4), Status Pekerjaan Kepala Rumah Tangga (X5), Status Kepemilikan Rumah (X6), Bahan Bangunan Atap (X7), Bahan Utama Dinding Rumah (X8), Bahan Utama Lantai Rumah (X9), Penggunaan Fasilitas BAB (X10), Sumber Air Minum (X11), Bahan Bakar Utama Memasak (X12), Luas lantai (X13), dan Wilayah (X14).

Langkah –langkah analisis data yang dilakukan dalam penelitian ini adalah sebagai berikut:

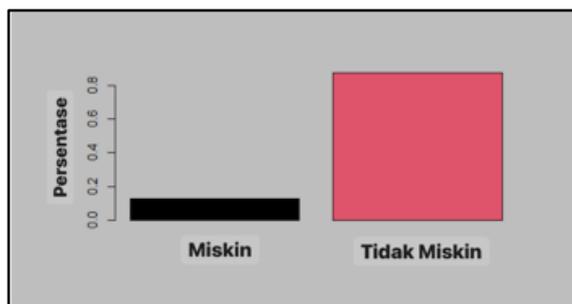
1. Pengklasifikasian status kemiskinan rumah tangga berdasarkan penghitungan garis kemiskinan per bulan per orang dengan jumlah anggota rumahtangga.
2. Eksplorasi data awal sehingga diperoleh gambaran umum dari data kemiskinan Provinsi Bengkulu, dengan menggunakan metode Khi-Kuadrat dan analisis Korelasi
3. Melakukan pengecekan data hilang (*missing data values*) dilakukan pengisian dengan data lain yang relevan, dan mengidentifikasi kondisi data (terkait proporsi antara rumah tangga miskin dan tidak miskin)
4. Membagi data menjadi data training dan testing: data training digunakan untuk pemodelan dan data testing digunakan untuk evaluasi performa klasifikasi. Data dibagi menjadi 75:25 dari total data yang tersedia.
5. Melakukan pemodelan klasifikasi dengan pohon klasifikasi tunggal (CART) dan pohon klasifikasi gabungan *Random Forest* serta *Xgboost* pada data training berdasarkan pembagian pada tahap 4.
6. Melakukan evaluasi performa klasifikasi menggunakan nilai-nilai pada confusion matrix hasil pemodelan pada tahap 5.
7. Melakukan perbaikan proporsi data pada kelas minoritas dengan menggunakan SMOTE, metode SMOTE diterapkan pada data training yang diperoleh dari tahap 4.
8. Melakukan pemodelan klasifikasi dengan

pohon klasifikasi tunggal (CART) dan pohon klasifikasi gabungan Random Forest serta Xgboost pada data training berdasarkan pembagian pada tahap 7.

9. Melakukan evaluasi performa klasifikasi menggunakan nilai-nilai pada confusion matrix hasil pemodelan tahap 8.
10. Memilih satu model terbaik berdasarkan nilai AUC, F-score, Sensitivity dan specificity
11. Mengidentifikasi variabel penting yang membangun model terbaik pada tahap 10.
12. Interpretasikan hubungan variabel penting dengan status rumah tangga.

HASIL DAN PEMBAHASAN

Pengklasifikasian rumah tangga yang tergolong miskin dan tidak miskin dilakukan berdasarkan garis kemiskinan Provinsi Bengkulu pada bulan Maret 2020. Total rumah tangga sampel adalah sebanyak 5730 rumah tangga, dengan rincian 721 rumah tangga terkategori miskin dan 5009 rumah tangga terkategori tidak miskin. Proporsi antara rumah tangga yang dikategorikan miskin jauh lebih kecil dibandingkan proporsi rumah tangga tidak miskin. Ketimpangan ini sering disebut sebagai kondisi data yang tidak seimbang (*unbalance dataset*).



Gambar 2. Sebaran Status Kemiskinan Rumah Tangga di Provinsi Bengkulu.

Sumber: Data SUSENAS 2020 yang diolah.
 Sumber: Data SUSENAS 2020 yang diolah

Terdapat 14 variabel penjelas yang digunakan pada penelitian ini yaitu Jenis Kelamin Kepala Rumah Tangga (X1), Usia Kepala Rumah Tangga (X2), Banyaknya Anggota Rumah Tangga (X3), Ijazah Tertinggi Kepala Rumah Tangga (X4),

Status Pekerjaan Kepala Rumah Tangga (X5), Status Kepemilikan Rumah (X6), Bahan Bangunan Atap (X7), Bahan Utama Dinding Rumah (X8), Bahan Utama Lantai Rumah (X9), Penggunaan Fasilitas BAB (X10), Sumber Air Minum (X11), Bahan Bakar Utama Memasak (X12), Luas lantai (X13), dan Wilayah (X14). Untuk menentukan variabel prediktor yang memiliki hubungan dengan variabel respon yang digunakan yaitu jika variabel prediktor merupakan variabel numerik maka digunakan uji t-student, sedangkan untuk variabel kategorik digunakan uji Khi-kuadrat.

Tabel 2. Hasil pengujian Khi-Kuadrat/Fisher

Variabel	Uji khi-kuadrat
Jenis Kelamin	Signifikan
Pendidikan	Signifikan
kepemilikan rumah	Signifikan
jenis atap	Signifikan
jenis lantai	Signifikan
sumber air minum	Signifikan
bahan bakar	Signifikan
fasilitas BAB	Signifikan
pekerjaan KRT	Tidak Signifikan
Wilayah	Signifikan

Sumber: Data SUSENAS 2020 yang diolah

Tabel 3. Hasil pengujian t-student

Variabel	Uji-t
Luas lantai	Signifikan
Usia	Signifikan
Jumlah anggota rumah tangga	Signifikan

Sumber: Data SUSENAS 2020 yang diolah

Pemodelan Klasifikasi Kemiskinan

a. Model menggunakan data awal

Analisis klasifikasi kemiskinan dilakukan menggunakan model klasifikasi pohon tunggal dan gabungan. Sebelum dilakukan analisis, dataset dibagi menjadi data *training* dan data *testing*. Data *training* digunakan untuk membangun model pembelajaran, kemudian hasil model akan dievaluasi menggunakan data *testing*. Pembagian dataset dilakukan secara acak dengan proporsi 75%:25%. Berikut adalah gambaran pembagian data *training* dan *testing*:

Tabel 4. Pembagian Data

	Total	Miskin	Tidak Miskin
Data Training	4296	540	3756
Data Testing	1434	181	1253

Sumber: Data SUSENAS 2020 yang diolah

Pemodelan menggunakan metode klasifikasi CART dilakukan dengan terlebih dahulu mencari parameter optimal yaitu *cost complexity*, kedalaman pohon dan minimal observasi dalam setiap simpul pohon. Berikut rincian beberapa kombinasi parameter yang menghasilkan performa klasifikasi paling baik:

Tabel 5. Kombinasi Parameter Optimal model CART

No	Cost	Kedalaman	Minsplit	AUC
1	2.5×10^{-5}	13	22	0.751
2	$8,2 \times 10^{-4}$	14	14	0.747
3	1.87×10^{-6}	7	13	0.747

Sumber: Data SUSENAS 2020 yang diolah

Berdasarkan Tabel 5. dapat dilihat bahwa terdapat kombinasi parameter yang memberikan nilai AUC paling tinggi. Selanjutnya akan dilakukan pemodelan klasifikasi CART dengan menggunakan *cost* sebesar 2.5×10^{-5} , kedalaman pohon sebesar 13 dan minimal observasi dalam simpul sebesar 22. Jika ditinjau dari kedalaman pohon, model yang dihasilkan cukup kompleks.

Sama halnya dengan pemodelan CART, pada *Random Forest* juga terdapat beberapa parameter yang harus ditetapkan diawal pemodelan. parameter tersebut antara lain adalah *mtry* (jumlah variabel prediktor yang diambil secara acak dari total variabel yang tersedia), *ntree* (banyak pohon yang dibentuk) dan *minspl* (minimal observasi pada setiap simpul). Berikut adalah beberapa kombinasi dari parameter dalam pemodelan random forest yang memberikan performa paling baik:

Tabel 6. Kombinasi Parameter Optimal model *Random Forest*

No	Mtry	Ntree	Minsplit	AUC
1	12	1270	40	0.796
2	12	691	36	0.794
3	12	1657	31	0.793

Sumber: Data SUSENAS 2020 yang diolah

Berdasarkan Tabel 6 diperoleh model *random forest* terbaik dihasilkan jika pemodelan dilakukan menggunakan parameter *mtry* sebesar 12, *ntree* sebesar 1270 dan *minspl* sebesar 40. *Ntree* sebesar 1270 mengindikasikan terdapat 1270 pohon klasifikasi tunggal yang terbentuk pada pemodelan.

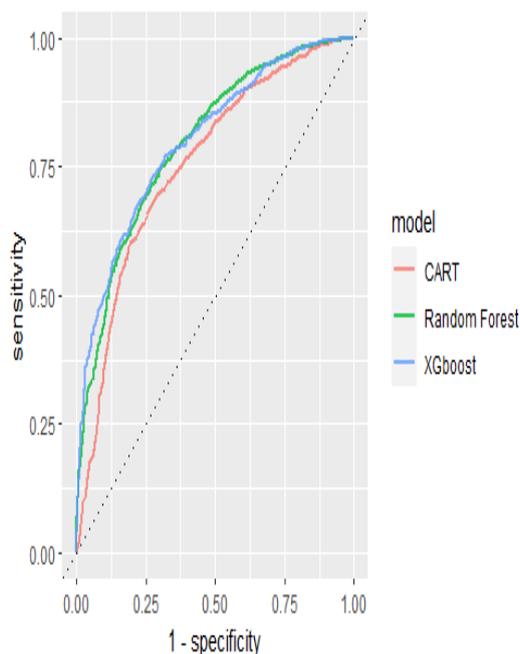
Pemodelan klasifikasi menggunakan *Xgboost* memiliki parameter yang mirip dengan *Random Forest* yaitu *mtry*, *ntree* dan *minspl*. Tabel 6 melampirkan kombinasi terbaik dari parameter ini, dengan nilai AUC paling tinggi jika pemodelan *Xgboost* dilakukan menggunakan *mtry* sebesar 3, *ntree* sebesar 413 dan *minspl* sebesar 24.

Tabel 7. Kombinasi Parameter Optimal model XGboost

No	Mtry	Ntree	Minsplit	AUC
1	3	413	24	0.800
2	3	864	19	0.791
3	3	1359	35	0.790

Sumber: Data SUSENAS 2020 yang diolah

Gambar 3 menyajikan perbandingan performa klasifikasi CART, *Random Forest* dan *Xgboost* pada data yang tidak seimbang (proporsi rumah tangga miskin jauh lebih sedikit). Dapat dilihat bahwa pemodelan dengan *Xgboost* berada paling atas, meskipun memiliki nilai yang tidak terlalu berbeda dengan *Random Forest*. Dari ketiga model, pemodelan dengan pohon klasifikasi tunggal (CART) memiliki performa klasifikasi yang paling rendah.



Gambar 3. Kurva ROC CART, *Random Forest* dan *Xgboost*

Sumber: Data SUSENAS 2020 yang diolah

Tabel 8. Evaluasi Model Klasifikasi

Model	F-score	A	B	AUC
CART	0.177	0.122	0.964	0.749
Random Forest	0.148	0.088	0.985	0.804
XGboost	0.261	0.182	0.969	0.801

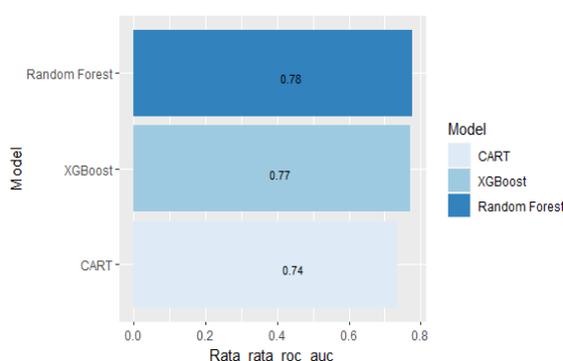
Sumber: Data SUSENAS 2020 yang diolah

dimana A ialah Sensitivity dan B ialah *Specivicity*. Tabel 8. Merupakan hasil evaluasi setiap pemodelan klasifikasi yang digunakan dalam penelitian ini. performa klasifikasi dinilai dari hasil prediksi pada data testing. Dikarenakan data memiliki permasalahan ketidakseimbangan data, maka performa klasifikasi dapat dibandingkan dengan melihat nilai F-score. Nilai F-score tertinggi diperoleh pada pemodelan *Xgboost*, namun pemodelan *Xgboost* tidak cukup baik dalam mengatasi permasalahan ketidakseimbangan data. Hal ini dapat dilihat dari nilai sensitivity sebesar 0.182, yang berarti metode ini hanya mampu mengklasifikasikan rumah tangga miskin secara tepat sebesar 18.2%.

Sehingga perlu penanganan permasalahan ini pada level data yaitu dengan metode SMOTE.

b. Model menggunakan data setelah SMOTE

Pemodelan CART pada data setelah SMOTE memiliki performa paling baik saat *cost_complexity* sebesar $2,74 \times 10^{-7}$, kedalaman pohon sebesar 6 dan *minspl* 25. Pemodelan Random Forest memiliki performa paling baik saat *n*tree sebanyak 1270, *m*try sebesar 12 dan *minspl* sebesar 25. Sedangkan pada *Xgboost*, performa paling baik saat *n*tree sebanyak 413, *m*try sebesar 3 dan *minspl* sebesar 24. Gambar 4 menyajikan perbandingan rata-rata nilai AUC yang diperoleh pada setiap model.



Gambar 4. Rata-rata AUC pada Setiap Model

Sumber: Data SUSENAS 2020 yang diolah

Selanjutnya, setiap pemodelan yang dihasilkan dievaluasi performa klasifikasinya dengan melihat nilai F-score, *sensitivity*, *specificity* dan AUC yang disajikan dalam Tabel 9. Evaluasi dilakukan dengan menerapkan setiap model pada data testing.

Tabel 9. Evaluasi Model Klasifikasi

Model	F-score	A	B	AUC
CART	0.344	0.569	0.749	0.736
Random Forest	0.365	0.409	0.879	0.791
XGboost	0.350	0.376	0.888	0.779

Sumber: Data SUSENAS 2020 yang diolah

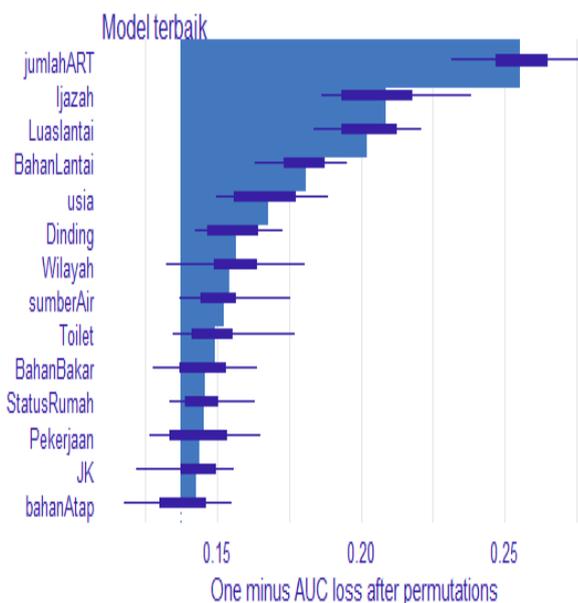
dimana A ialah Sensitivity dan B ialah *Specivicity*

Berdasarkan Tabel 9 dapat dilihat bahwa model Random Forest memiliki nilai AUC paling tinggi dan nilai sensitivity tidak terlalu rendah yaitu 0.409 dan *specivicity* 0.879 Sehingga dalam interpretasi model akan digunakan Random Forest sebagai

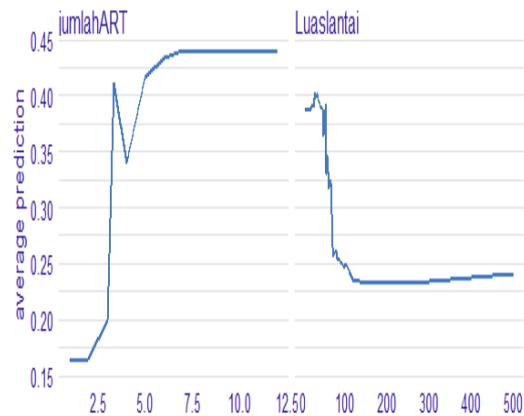
identifikasi karakteristik rumah tangga miskin dan tidak miskin.

Variable Importance

Variable Importance digunakan untuk melihat tingkat kepentingan setiap variabel yang digunakan di dalam pemodelan yang terpilih pada tahapan sebelumnya. Dari 14 variabel prediktor, terdapat tiga variabel prediktor yang memiliki kontribusi paling tinggi dalam mengklasifikasikan rumah tangga miskin dan tidak miskin, yaitu : jumlah anggota rumah tangga, ijazah terakhir yang dimiliki kepala keluarga dan luas lantai rumah tempat tinggal (Gambar 5). Hubungan antara variabel tersebut terhadap rumah tangga miskin dapat dilihat dari Gambar 6. Hubungan antara jumlah anggota rumah tangga dengan rumah tangga miskin adalah positif, yang mengindikasikan bahwa semakin banyak anggota rumah tangga akan semakin besar peluang rumah tangga tersebut masuk kedalam kategori miskin. Sedangkan luas lantai rumah memiliki hubungan negatif yang berarti semakin sempit rumah yang ditempati maka semakin besar peluang rumah tangga tersebut terkategori miskin.



Gambar 5. *Variable Importance* Model Random Forest pada data SMOTE
Sumber: Data SUSENAS 2020 yang diolah



Gambar 6. Partial Dependence Plot Variabel Jumlah ART dan Luas Lantai.

Sumber: Data yang diolah

KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan yang diuraikan pada bab sebelumnya, dapat diambil kesimpulan bahwa: Pemodelan dengan pohon klasifikasi pohon gabungan dalam hal ini Random Forest dan Xgboost tidak dapat memberikan performa yang baik pada data yang tidak seimbang. Penerapan metode SMOTE pada data dapat meningkatkan ketepatan klasifikasi pada kelas minoritas (rumah tangga miskin), namun belum maksimal. Terdapat tiga variabel yang paling berkontribusi besar dalam pengklasifikasian dengan metode SMOTE+Random Forest yaitu jumlah anggota rumah tangga, ijazah terakhir kepala rumah tangga dan luas lantai rumah. Temuan ini diharapkan dapat digunakan sebagai dasar analisis pengambilan keputusan, khususnya untuk mengidentifikasi rumahtangga miskin, sehingga program-program bantuan untuk penganggulangan kemiskinan diharapkan lebih tepat sasaran di masa mendatang.

Pada penelitian selanjutnya, metode penanggulangan ketidakseimbangan data lainnya sebaiknya juga dilakukan, sehingga ketepatan klasifikasi lebih tinggi pada kelas minoritas.

DAFTAR PUSTAKA

Badan Pusat Statistik (BPS). 2020. Diakses pada 29 Maret 2021 melalui: <https://www.bps.go.id/pressrelease/2021/02/15>

/1851/persentase-penduduk-miskin-september-2020-naik-menjadi-10-19-persen.html

- Breiman L., Friedman J.H., Olshen R.A., Stone C.J. 1984. Classification and Regression Trees. New York: Chapman & Hall/CRC.
- Chawla N.V., Browyer K.W., Hall L.O., Kegelmeyer W.P. 2002. SMOTE : Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research. 16, 321-357.
- Chen, T., & Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
- Han, J., Pei, J., & Kamber, M. 2012. Data mining: concepts and techniques. Elsevier.
- Fawcett T. 2006. An introduction to ROC analysis. Pattern Recognition Letters. 27, 861–874.
- Ishartono dan Raharjo S.T. 2016. Sustainable Development Goals (SDGs) dan Pengetasan Kemiskinan. Share: Social Work Jurnal, Vol 6, 2
- Ispriyanti, D., Prahutama, A., dan Mustafid. 2019. Analisis Klasifikasi Kemiskinan Kota Semarang Menggunakan Algoritma QUEST. Statistika, Vol.7(1)
- Lewis, R. J. 2000. An introduction to classification and regression tree (CART) analysis. In Annual meeting of the society for academic emergency medicine in San Francisco, California (Vol. 14).
- Liaw A., Wiener M. 2002. Classification and Regression. R News, 2,18-22
- Sartono, B. dan Syafitri., T. U. 2010. Metode Pohon Gabungan: Solusi Pilihan Untuk Mengatasi Kelemahan Pohon Regresi Dan Klasifikasi Tunggal. Forum Statistika dan Komputasi, 1-7
- Sutton C. D. 2005. Classification and Regression Trees, Bagging, and Boosting. Handbook of Statistics, 24, 303-329.
- Wezel M.V., Potharst R. 2007. Improved Customer Choice Predictions using Ensemble Methods. European Journal of Operational Research, 181, 43

PEMODELAN PREVALENSI PENYAKIT KUSTA DI JAWA TIMUR DENGAN PENDEKATAN *GEOGRAPHICALLY WEIGHTED REGRESSION*

Christopher Andreas¹, Horidah Horidah², Rizza Sulistiana³, Dhyana Venosia⁴, dan Nur
Chamidah⁵

1,2,3,4,5Departemen Matematika Universitas Gadjah Mada, Yogyakarta-Indonesia
e-mail: 1lagubu2014@gmail.com, 2dedirosadi@gadjahmada.edu, 3rachmanstat@ugm.ac.id

Received: January 5, 2022; Revised: February 23, 2022; Accepted: May 15, 2022; Published: June 30, 2022

Abstrak

Indonesia merupakan penyumbang kasus baru penyakit kusta tiga terbesar lingkup dunia, dengan Jawa Timur sebagai provinsi yang berkontribusi tertinggi. Hal ini membuat pemerintah menetapkan target rencana aksi guna mengeliminasi kusta di Indonesia. Penyakit kusta tergolong penyakit menular sehingga penyebaran penyakit ini dipengaruhi faktor lingkungan geografis. Oleh sebab itu, digunakan metode *Geographically Weighted Regression* (GWR) karena dapat mengestimasi nilai parameter disetiap titik pengamatan melalui aspek spasial. Pada pemodelan prevalensi penyakit kusta di Jawa Timur, dengan persentase rumah tangga yang menggunakan fasilitas mandi, cuci, kakus bersama (X_1), persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak (X_2), dan jumlah penduduk miskin (X_3) memberi pengaruh signifikan terhadap prevalensi penyakit kusta di Jawa Timur, dengan sebaran kabupaten dan kota yang dipengaruhi secara signifikan masing – masing sebanyak 26,32%, 18,42%, dan 39,47%. Selain itu, hasil pemodelan GWR menunjukkan akurasi dan kebaikan model yang lebih baik daripada model regresi global yang ditinjau berdasarkan nilai *Symmetric Mean Absolute Percentage Error* (sMAPE) dan *Akaike Information Criterion* (AIC), yang masing-masing sebesar 12,14% dan -73,124. Pada penelitian ini, diketahui bahwa prevalensi penyakit kusta tertinggi di Jawa Timur terjadi di Kabupaten Sumenep dengan nilai prevalensi sebesar 2,06. Untuk mencapai target eliminasi kusta yang ditetapkan pemerintah, maka Pemerintah Kabupaten Sumenep dapat menurunkan nilai persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak (X_2) hingga mencapai 12,01% dan jumlah penduduk miskin (X_3) hingga mencapai 13,215 ribu jiwa. Pemerintah Jawa Timur perlu fokus untuk menanggulangi variabel-variabel yang signifikan disetiap wilayah masing-masing kabupaten atau kota, supaya bisa menurunkan prevalensi kusta sampai target yang ditetapkan oleh pemerintah. Hasil penelitian dapat digunakan sebagai landasan dalam merumuskan rekomendasi kebijakan di setiap kabupaten dan kota di Jawa Timur guna mencapai target eliminasi kusta yang ditetapkan pemerintah.

Kata Kunci : Analisis Spasial, Eliminasi Kusta, Metode GWR, Prevalensi Penyakit Kusta, Regresi Global.

Abstract

Indonesia is the largest contributor to new cases of leprosy in the world, with East Java as the province with the highest contribution. This has made the government set a target for an action plan to eliminate leprosy in Indonesia. Leprosy is classified as an infectious disease so that the spread of this disease is influenced by geographical environmental factors. Therefore, the Geographically Weighted Regression (GWR) method is used because it can estimate the parameter value at each observation point through the spatial aspect. In modeling the prevalence of leprosy in East Java, with the percentage of households using shared bathing, washing, and latrine facilities (X_1), the percentage of households that do not have access to sanitation feasible (X_2), and the number of poor people (X_3) had a significant influence on the prevalence of leprosy in East Java, with the distribution of districts and cities being significantly affected respectively by 26.32%, 18.42 %, and 39.47%. In addition, the results of the GWR modeling show that the accuracy and goodness of the model are better than the global regression model reviewed based on the Symmetric Mean Absolute Percentage Error (sMAPE) and Akaike Information Criterion (AIC) values, which are 12,14% and -73,124, respectively. In this study, it was found that the highest prevalence of leprosy in East Java occurred in Sumenep Regency with a prevalence value of 2,06. To achieve the leprosy elimination target set by the government, the Sumenep Regency Government can reduce the percentage of households that do not have access to proper sanitation (X_2) to 12,01% and the number of poor people (X_3) to reach 13,215 thousand inhabitants. The East Java government needs to focus on tackling significant variables in each district or city, in order to reduce the prevalence of leprosy to the target set by the government. The results of the study can be used as a basis for formulating policy recommendations in every district and city in East Java to achieve the leprosy elimination target set by the government.

Keyword : Spatial Analysis, Elimination of Leprosy, GWR Method, Prevalence of Leprosy, Global Regression.

PENDAHULUAN

Berdasarkan *Sustainable Development Goals* (SDGs), penyakit menular atau *infectious disease* merupakan salah satu fokus kajian utama di bidang kesehatan. Salah satu penyakit menular tersebut adalah penyakit kusta. Dalam buku rencana aksi SDGs yang dikeluarkan oleh Bappenas pada tahun 2020, salah satu target rencana aksi SDGs bidang kesehatan di Indonesia adalah mengakhiri epidemi *Acquired Immune Deficiency Syndrome* yang biasa disebut dengan AIDS, Tuberkulosis yang biasa disebut dengan TBC, malaria, dan penyakit tropis seperti filariasis dan kusta (Bappenas, 2020). Salah satu penyakit kulit menular ialah penyakit kusta. Penyakit ini dipicu oleh bakteri *Mycobacterium Leprae*. Penyakit Kusta ini biasa dikenal dengan sebutan “*The Great Imitator Disease*”. Hal ini, karena keberadaan kusta sulit untuk teridentifikasi melalui gejala yang ditimbulkan (Tami, 2019). Penyakit kusta pada umumnya menyerang bagian kulit, mukosa pada saluran pernapasan, mata, dan saraf tepi. Penanganan terhadap terjadinya kasus penyakit kusta yang belum memadai akan menyebabkan penyakit kusta berkembang. Penderita penyakit kusta dapat mengalami masalah pada bagian kulit, mata, saraf, dan anggota gerak tubuh.

Negara Indonesia ialah salah satu dari sekian negara berkembang yang berpotensi terindikasi menjadi negara penyumbang terhadap kasus baru pada penyakit kusta urutan ke-3 terbesar lingkup dunia terkait jumlah kasus yang diciptakan sebesar 8% terhadap kasus kusta yang ada di dunia dimana 9.061 kasus baru kusta ditemukan (Kemenkes, 2021a). Fenomena penyakit kusta tahun 2020 teridentifikasi penemuan kasus baru sebanyak 11.173 di Indonesia dengan 86% diantaranya adalah kusta bertipe *Multibasiler* atau MB (Kemenkes, 2021b). Pendeteksian terjadinya kasus baru penyakit kusta diperlukan untuk mengendalikan peningkatan kasus baru penyakit kusta yang terjadi. Dalam meninjau terkait keberhasilan pendeteksian kasus penyakit kusta baru dapat dilakukan dengan penerapan penggunaan indikator terhadap angka kecacatan pada tingkat-2. Sebesar 1,18 dari 1.000.000 jumlah penduduk, merupakan angka kecacatan pada tingkat ke-2 yang

ditetapkan untuk tahun 2020, dimana angka yang ditetapkan akan berubah-ubah setiap tahunnya dan cenderung menurun. Hal ini, mengindikasikan bahwa terjadinya keterlambatan penemuan kasus penyakit kusta baru berpotensi dapat dicegah (Kemenkes, 2021a). Meskipun demikian, penanggulangan kusta di Indonesia perlu terus ditingkatkan.

Beberapa strategi yang diperlukan untuk memenuhi target penanggulangan kusta di Indonesia telah tertuang pada UU Nomor 11 Tahun 2019 tentang penanggulangan kusta pada Bab 2 Pasal 2 Ayat 1 & 3 yang berisi tentang target eliminasi kusta di Indonesia. Perwujudan terhadap penanggulangan terkait kasus penyakit kusta baru tersebut merupakan suatu tanggung jawab yang dimiliki oleh seluruh golongan masyarakat seperti tercantum pada UU Nomor 11 Tahun 2019 Pasal 5 Ayat (1).

Indonesia adalah sebuah negara kesatuan dengan 34 provinsi, dimana Jawa Timur merupakan satu diantara provinsi lainnya yang terindikasi memiliki beban terhadap penyakit kusta tertinggi. Hal ini, membuat provinsi Jawa Timur menjadi provinsi dengan penyumbang kontribusi yang besar mengenai jumlah pada kasus penyakit kusta yang ada di negara Indonesia (Ritianty dkk., 2020). Maka, Jawa Timur terindikasi sebagai provinsi dengan jumlah banyaknya penyakit kusta tertinggi yang berada di Indonesia (Kominfo Jatim, 2020). Selama tahun 2019 hingga 2020, pengendalian kasus kusta di Jawa Timur mengalami peningkatan yang ditinjau berdasarkan mengecilnya penemuan kusta dari 8,06 menjadi 4,19 per 1.000.000. Guna mewujudkan eliminasi kusta di Indonesia, khususnya pada Provinsi Jawa Timur, pengoptimalan faktor yang diasumsikan berpengaruh signifikan pada penyakit kusta harus dilakukan. Maka, dilakukan suatu penelitian guna membentuk pemodelan statistika terkait prevalensi kasus penyakit kusta di Jawa Timur sehingga faktor yang mempengaruhi jumlah kasus pada penyakit kusta dapat diidentifikasi dengan jelas.

Penelitian terkait telah dilakukan oleh Pertiwi, dkk. (2020), yang menyimpulkan bahwa persentase rumah sehat memberikan pengaruh terkait jumlah terkait kasus pada penyakit kusta untuk kabupaten atau kota yang

berada di Jawa Timur. Kemudian, menurut Zuhdan dkk. (2017), kondisi mengenai ekonomi keluarga dengan pendapatan yang kurang serta kebersihan yang buruk individu mempengaruhi peningkatan jumlah kasus penyakit kusta. Maka, dapat diduga faktor terkait yang dapat mempengaruhi terjadinya kasus penyakit kusta antara lain terkait aspek kebersihan dan ekonomi. Oleh sebab itu, penduga variabel yang digunakan dalam memodelkan terkait kasus mengenai penyakit kusta untuk kabupaten atau kota yang berada di Jawa Timur antara lain, fasilitas Mandi, Cuci, Kakus Bersama (MCK Bersama), sanitasi belum layak, dan penduduk miskin. Variabel yang digunakan dalam penelitian ini, memiliki kaitan yang erat dengan fokus kajian bersama di bidang kesehatan yang tertuang dalam SDGs yaitu *water-borne disease* dan *water, sanitation, and hygiene*.

Kemudian, dalam upaya memetakan penyakit kusta untuk setiap kabupaten atau kota yang berada di Jawa Timur digunakan pendekatan spasial. Pada pendekatan spasial menggunakan penaksiran dengan titik pengamatan, sehingga memperoleh nilai parameter yang berbeda-beda (Widayaka dkk, 2016). Dalam hal ini, pada model regresi spasial terdapat *spatial dependence* dan *spatial heterogeneity* sebagai efek spasialnya. Salah satu pendekatan spasial yakni metode *Geographically Weighted Regression* yang biasa disebut dengan GWR. Pemilihan metode GWR dilakukan untuk memperoleh nilai parameter setiap titik pengamatan pada data yang bersifat kontinu melalui aspek spasial (Widayaka dkk, 2016). Metode GWR, merupakan suatu metode yang dapat digunakan dalam melakukan pemodelan variabel independen terhadap variabel dependen dengan melibatkan unsur wilayah. Keuntungan dari model GWR yaitu dapat melakukan suatu pemodelan berdasarkan faktor wilayah yang tidak dapat dilakukan dalam pemodelan regresi klasik. Dziauddin & Idris (2017), mengemukakan bahwa pada model GWR memiliki sebuah kemampuan dalam menghasilkan hasil yang akurat karena mampu memberikan perbedaan atribut pada efek spasial di setiap wilayah berdasarkan daerah geografisnya terkait nilai suatu properti

residensial. Kemudian, estimasi parameter yang ada pada model melalui GWR digunakan *Weighted Least Square* yang biasa disebut dengan WLS dengan menggunakan pembobot fungsi *kernel gaussian*. Maka, dapat disimpulkan bahwa pemodelan dengan menggunakan model GWR lebih akurat daripada model yang diperoleh melalui regresi linier karena model GWR digunakan dengan tujuan untuk memperoleh titik pengamatan dengan perbedaan hasil nilai parameter yang tidak dapat dilakukan jika menggunakan regresi global (Widayaka dkk, 2016). Selain itu, terkait variabel yang signifikan untuk setiap daerah pada model GWR adalah berbeda.

Berdasarkan uraian diatas, dalam hal ini peneliti menduga bahwa persebaran kasus terkait penyakit kusta disebabkan oleh kondisi geografis yang terjadi pada kabupaten atau kota yang berada di Jawa Timur. Fenomena ini, diduga karena adanya faktor kedekatan wilayah yang memberikan pengaruh sehingga penggunaan metode GWR telah dianggap sesuai. Pada kasus ini, diharapkan dengan menggunakan model GWR, diperoleh nilai parameter yang berbeda untuk setiap kabupaten atau kota yang berada di Jawa Timur sehingga hasil estimasi prediksi yang diperoleh sesuai dengan realita di lapangan.

Menurut Muharry (2014), kebersihan dan ekonomi merupakan salah satu faktor risiko yang mempengaruhi terjadinya penyakit kusta. Analisis pemodelan prevalensi kasus penyakit kusta yang dipengaruhi oleh variabel yang meliputi aspek kebersihan dan ekonomi dengan metode GWR menjadi kebaruan penelitian ini. Pada penelitian sebelumnya, Muharry (2014) menggunakan desain studi kasus kontrol dengan kesimpulan yang diperoleh yaitu kondisi ekonomi keluarga rendah dan kebersihan perorangan buruk mempengaruhi terjadinya penyakit kusta. Kemudian, dengan metode yang sama Zuhdan dkk (2017), menyimpulkan bahwa faktor yang berpengaruh terhadap kejadian penyakit kusta antara lain tingkat pendidikan rendah, lama kontak ≥ 1 tahun, status gizi buruk, kondisi ekonomi keluarga yang berpendapatan kurang, dan kebersihan perorangan yang buruk. Penelitian ini secara khusus juga memberikan

kontribusi berupa pemetaan wilayah terkait kasus penyakit kusta sehingga pemerintah Jawa Timur dapat mempersiapkan upaya *tracing* dengan lebih sigap agar kasus yang terjadi tidak meluas ke seluruh daerah, serta sebagai landasan dalam membuat kebijakan untuk mengeliminasi penyakit kusta khususnya di Jawa Timur, sesuai dengan tujuan dalam SDGs.

TINJAUAN PUSTAKA

Prevalensi Penyakit Kusta

Prevalensi penyakit kusta per 10.000 penduduk didefinisikan sebagai jumlah kasus kusta yang terdaftar per 10.000 penduduk pada wilayah dan kurun waktu tertentu (Ernawati dkk., 2016). Berdasarkan definisi tersebut, dapat diperoleh rumus untuk menghitung prevalensi penyakit kusta di suatu wilayah sebagai berikut.

$$\text{prevalensi} = \frac{\text{jumlah kasus kusta}}{\text{jumlah penduduk}} \times 10.000$$

Pemerintah pusat menetapkan target eliminasi penyakit kusta berdasarkan Peraturan Menteri Kesehatan Republik Indonesia Nomor 11 Tahun 2019 tentang Penanggulangan Kusta Pasal 2 Ayat 1 (Kemenkes, 2019). Indikator untuk mencapai target eliminasi kusta sesuai dengan yang ditetapkan oleh pemerintah pusat adalah kurang dari 0,0001 atau dengan kata lain apabila prevalensi kasus penyakit kusta di suatu daerah melebihi target tersebut maka daerah tersebut belum memenuhi target penanggulangan kusta.

Kategori beban kusta dapat diklasifikasikan menjadi 2 diantaranya beban kusta tinggi (*high burden*) dan beban kusta rendah (*low burden*) (Indayani dkk., 2017). Suatu daerah dikatakan memiliki beban kusta tinggi apabila jumlah kasus baru kusta lebih dari 0,0001 penduduk dan atau jumlah kasus baru lebih dari 1.000 kasus. Sementara itu, daerah yang memiliki beban kusta rendah adalah daerah yang memiliki jumlah kasus baru kusta kurang dari 0,0001 penduduk.

Faktor Penyebab Penyakit Kusta

Persebaran penyakit kusta di Jawa Timur dipengaruhi oleh kondisi geografis. Hal ini disebabkan adanya keterkaitan antara daerah yang satu dengan daerah yang lainnya. Selain

itu, terdapat variabel pada aspek kebersihan dan ekonomi yang diduga mempengaruhi persebaran penyakit kusta.

Fasilitas Mandi, Cuci, Kakus Bersama (MCK Bersama)

Menurut BPS, sejak tahun 2019 penggunaan MCK bersama merupakan salah satu indikator rumah tangga yang memiliki akses terhadap sanitasi layak yang digunakan secara bersama pada Anggota Rumah Tangga (ART). MCK bersama memiliki hubungan yang erat dengan rumah dan lingkungan, sehingga dapat mempengaruhi kesehatan. Kebersihan MCK yang digunakan bersama menentukan keberlangsungan berkembangbiaknya berbagai macam bakteri, salah satunya bakteri kusta. Kondisi MCK bersama yang sehat, dapat mencegah penularan dan kemunculan penyakit, terutama penyakit kusta yang memiliki keterkaitan yang erat dengan kebersihan MCK (Putri dkk., 2021).

Sanitasi Belum Layak

Menurut Rasyidah (2019), akses terhadap sanitasi yang belum layak menyebabkan timbulnya berbagai macam penyakit yang disebabkan oleh faktor kebersihan. Salah satu penyakit yang berpotensi dapat ditimbulkan dari pengaruh sanitasi yang belum layak adalah penyakit kusta. Sanitasi belum layak sebagaimana yang dimaksud antara lain akses tidak memadainya akses air bersih, tidak sehatnya fasilitas jamban, dan perilaku mencuci tangan yang higienis (Olo dkk., 2021). Hal ini disebabkan oleh air yang merupakan perantara penularan penyakit kusta sekaligus bertindak sebagai reservoir untuk *Mycobacterium leprae* (Siswanti & Wijayanti, 2018).

Penduduk Miskin

6. Konsep penduduk miskin menurut Badan Pusat Statistik (2020) adalah penduduk dengan rata-rata pengeluaran per kapita tiap bulan berada di bawah garis kemiskinan. Garis kemiskinan sebagaimana yang dimaksud meliputi Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non-Makanan (GKNM). Dalam hal ini, tingkat ekonomi dapat mempengaruhi derajat kesehatan karena ketidakmampuan dalam memenuhi kebutuhan sehari-hari sehingga kebutuhan akan kesehatan belum tentu terjamin. Hal ini membuktikan

bahwa penderita kusta didominasi oleh kelompok dengan sosial ekonomi rendah atau dapat dikatakan miskin (Muntasir dkk., 2018).

Analisis Regresi Linier Berganda

Untuk mengetahui hubungan fungsional antara beberapa variabel prediktor dengan satu variabel respon maka digunakan analisis regresi (Paolella, 2019). Secara umum, model regresi linier berganda sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{ip-1} + \varepsilon_i$$

Dengan Y_i adalah variabel respon pengamatan ke- i , β_0 adalah konstanta atau intersep, β_1 adalah parameter regresi variabel prediktor 1 pengamatan ke- i , X_{i1} adalah variabel prediktor 1 pengamatan ke- i , p adalah jumlah prediktor dan ε_i adalah galat pengamatan ke- i .

Pada analisis regresi, terdapat asumsi yang harus terpenuhi untuk mendapatkan estimator tidak bias dan memiliki variansi minimum dengan metode kuadrat terkecil. Asumsi tersebut yaitu galat harus berdistribusi normal, tidak terjadi kasus multikolinieritas dan heteroskedastisitas, serta tidak ada autokorelasi dalam galat apabila data berupa data *time series* (Olive, 2017; Andreas, dkk., 2021).

Asumsi Spasial Dalam Analisis Data Spasial

Pada analisis data spasial, terdapat asumsi yang harus terpenuhi yaitu autokorelasi spasial dan heterogenitas spasial. Autokorelasi spasial terjadi karena adanya ketergantungan antar lokasi pengamatan. Autokorelasi spasial dapat diketahui dengan menggunakan metode *Moran's Index* (Wuryandari dkk., 2014). Statistik uji untuk asumsi ini adalah sebagai berikut:

$$Z(I) = \frac{I - E(I)}{\sqrt{\text{var}(I)}}$$

$$\text{dengan } I = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

$$E(I) = -\frac{1}{n-1}; \text{var}(I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{(n^2 - 1) S_0^2} - [E(I)]^2$$

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}; S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2;$$

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

dengan x_i adalah data lokasi ke- i , x_j adalah data lokasi ke- j , \bar{x} adalah rata-rata data, w_{ij} adalah matriks pembobot, I adalah nilai indeks moran, $E(I)$ adalah rata-rata indeks moran,

dan $\text{var}(I)$ adalah variansi dari indeks moran. Asumsi autokorelasi spasial dapat terpenuhi dalam taraf signifikansi α apabila $|Z(I)| > Z_{\frac{\alpha}{2}}$ atau nilai *p-value* $< \alpha$.

Heterogenitas spasial terjadi karena adanya keragaman antar lokasi observasi. Dalam hal ini, uji *Breusch-Pagan* digunakan untuk mengetahui adanya heterogenitas spasial (Fatati dkk., 2017). Statistik uji untuk asumsi ini adalah sebagai berikut:

$$BP = \frac{1}{2} [f^T Z (Z^T Z)^{-1} Z^T f] \sim \chi_p^2$$

dengan $f_i = \left(\frac{\varepsilon_i^2}{\sigma^2} - 1 \right)$ dan Z adalah matriks dengan ukuran $n \times (p + 1)$. Matriks tersebut berisi vektor yang telah dinormalkan (z) untuk setiap pengamatan. Sementara ε_i^2 adalah galat pengamatan ke- i , σ^2 adalah variansi galat, dan p adalah jumlah prediktor.

Asumsi heterogenitas spasial dapat terpenuhi dalam taraf signifikansi α apabila $BP > \chi^2_{\alpha(p-1)}$ atau nilai *p-value* $< \alpha$.

Geographically Weighted Regression (GWR)

7. Pada dasarnya, *Geographically Weighted Regression* (GWR) merupakan metode yang didasarkan pada model regresi. Setiap lokasi memiliki nilai parameter regresi yang berbeda dalam metode GWR (Chamidah dkk., 2014). Hal ini dikarenakan parameter dihitung di setiap lokasi observasi. Variabel prediktor yang koefisien regresinya bergantung pada lokasi pengamatan dapat digunakan untuk memprediksi variabel respon pada model GWR (Sediono, dkk., 2022). Model GWR dapat dituliskan sebagai berikut (Permai dkk., 2021):

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

dengan y_i adalah variabel respon pada lokasi ke- i , x_{ik} adalah variabel prediktor pada lokasi ke- i , (u_i, v_i) adalah koordinat lokasi ke- i , $\beta_0(u_i, v_i)$ adalah intersep yang bervariasi secara geografis, $\beta_k(u_i, v_i)$ adalah koefisien untuk variabel bebas k di lokasi ke- i dan ε_i adalah galat di lokasi ke- i .

Estimasi Parameter Model GWR

Estimasi parameter lokal pada model GWR dilakukan dengan cara memberikan pembobot yang berbeda di setiap lokasi observasi.

Pembobot ini disebut dengan metode WLS atau *Weighted Least Square* (Magri dan Ispriyanti, 2017). Estimator dari model GWR adalah sebagai berikut (Lu dkk., 2014):

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y}$$

dengan $\mathbf{W}(u_i, v_i)$ adalah diagonal matriks pembobot untuk lokasi (u_i, v_i) , $\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}$ adalah matriks varian-kovarian terbobot secara geografis dan \mathbf{Y} adalah vektor dari nilai-nilai variabel prediktor.

Pembobot Model GWR

Fungsi pembobot model GWR digunakan untuk memberikan nilai dugaan parameter yang berbeda di setiap lokasi. Fungsi kernel dapat digunakan untuk menentukan besarnya pembobot di setiap lokasi (Kusnandar dkk., 2021). Menurut Pratiwi dkk. (2019), fungsi pembobot kernel tersebut antara lain:

1. Fungsi *Fixed Kernel*, yaitu pembobot untuk semua lokasi pengamatan yang memiliki nilai *bandwidth* yang sama, dengan uji statistik yang digunakan sebagai berikut:

$$8. W_j(u_i, v_i) = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{b} \right)^2 \right]$$

dengan d_{ij} adalah jarak antara lokasi ke- i dan lokasi ke- j yang diperoleh dari jarak *euclidian* $(d_{ij})^2 = (u_i - u_j)^2 + (v_i - v_j)^2$ dan b adalah parameter non negatif yang disebut *bandwidth*.

2. Fungsi *Adaptive Kernel*, yaitu pembobot untuk setiap lokasi pengamatan yang memiliki nilai *bandwidth* berbeda, dengan uji statistik yang digunakan sebagai berikut:

$$9. W_j(u_i, v_i) = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b} \right)^2 \right]^2 & ; \text{jika } d_{ij} < b \\ 0 & ; \text{lainnya} \end{cases}$$

Uji Kesesuaian Model GWR (*Goodness of Fit*)

Uji kesesuaian model GWR dilakukan untuk mendeteksi secara global apakah model GWR lebih baik daripada regresi global (Diastina dkk., 2019). Uji kesesuaian model GWR dilakukan dengan mengombinasikan uji regresi linier dengan model data spasial.

Pengujian Signifikansi Parameter Secara Parsial

Uji signifikansi parameter dari model GWR secara parsial dilakukan dengan tujuan untuk mengidentifikasi variabel prediktor apa saja

yang memiliki pengaruh signifikan terhadap variabel respon di setiap lokasi (Utami dkk., 2016).

Ukuran Kebaikan Model

Ukuran kebaikan model dilakukan dengan tujuan untuk membandingkan kebaikan dari model regresi secara global dengan model GWR yang mempertimbangkan unsur spasial. Ukuran tersebut diantaranya adalah koefisien determinasi, koefisien determinasi terkoreksi, dan *Akaike Information Criterion* (AIC).

Koefisien Determinasi

$$R^2 = 1 - \frac{JKG}{JKT}$$

dengan $JKG = \sum_{i=1}^n (y_i - \hat{y}_i)^2$; $JKT = \sum_{i=1}^n (y_i - \bar{y})^2$

Keterangan :

y_i : variabel respon pengamatan ke- i

\hat{y}_i : nilai dugaan variabel respon pengamatan ke- i

\bar{y} : nilai rata-rata variabel respon

Koefisien Determinasi Terkoreksi

$$10. R_{\alpha}^2 = 1 - \frac{JKG / (n-p)}{JKT / (n-1)}$$

Dengan n adalah jumlah pengamatan dan p adalah jumlah prediktor (Wohon, dkk., 2017). *Akaike Information Criterion* (AIC)

$$AIC = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n + tr(S)$$

dengan $\hat{\sigma} = \frac{JKG}{n}$ dan S adalah matriks proyeksi ukuran $n \times n$ dimana $\hat{y} = S\mathbf{y}$.

Bentuk matriks S sebagai berikut.

$$S = \begin{bmatrix} \mathbf{X}_1^T [\mathbf{X}^T \mathbf{W}(r_1, s_1) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(r_1, s_1) \\ \mathbf{X}_2^T [\mathbf{X}^T \mathbf{W}(r_2, s_2) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(r_2, s_2) \\ \vdots \\ \mathbf{X}_n^T [\mathbf{X}^T \mathbf{W}(r_n, s_n) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(r_n, s_n) \end{bmatrix}$$

dengan $\mathbf{W}(r_1, s_1)$ adalah matriks berukuran $n \times n$ (Lu, dkk., 2014).

Ukuran Akurasi Model

Ukuran akurasi model dilakukan dengan tujuan untuk mengukur keakuratan suatu model statistik dalam melakukan prediksi. Dalam penelitian ini, ukuran tersebut dihitung menggunakan sMAPE (*symmetric Mean Absolute Percentage Error*). Ukuran sMAPE digunakan karena terdapat beberapa data aktual yang mempunyai nilai nol dan dirumuskan sebagai berikut: (Crenata, dkk., 2012).

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(y_i + \hat{y}_i)}$$

METODOLOGI

Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder tahun 2020 yang diperoleh dari publikasi Badan Pusat Statistik dengan judul “STATISTIK INDONESIA 2021” (Badan Pusat Statistik, 2021).

Variabel Penelitian

Variabel yang digunakan meliputi variabel prediktor dan variabel respon. Secara rinci, terdapat 4 variabel yang terdiri dari 3 variabel prediktor dan 1 variabel respon. Variabel tersebut antara lain yaitu:

1. Prevalensi kasus penyakit kusta di Provinsi Jawa Timur (Y)
2. Persentase rumah tangga yang menggunakan fasilitas Mandi, Cuci, Kakus Bersama (MCK Bersama) di Provinsi Jawa Timur (X_1)
3. Persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak (X_2)
4. Jumlah penduduk miskin (X_3)

Selain itu, terdapat pula variabel geografis yaitu u_i berupa *latitude* (garis lintang) dan v_i berupa *longitude* (garis bujur) dari masing-masing lokasi.

Prosedur Analisis

11. Tahap-tahap yang dilakukan dalam analisis data yaitu antara lain:

1. Membuat deskripsi prevalensi penyakit kusta di Jawa Timur dan berbagai variabel yang mempengaruhinya berdasarkan peta tematik.
2. Membuat model statistika terkait prevalensi penyakit kusta di Jawa Timur yang diperoleh dengan metode GWR dengan tahapan antara lain:
 - a. Melakukan pengujian asumsi spasial pada data variabel respon yaitu sebagai berikut:
 - i. Melakukan uji Moran's I untuk menguji dependensi data, dengan kriteria jika $p\text{-value} < \alpha$ maka data memenuhi aspek spasial dependensi, berlaku juga sebaliknya.

- ii. Melakukan uji Breusch-Pagan untuk menguji heterogenitas data, dengan kriteria jika jika $p\text{-value} < \alpha$ maka data memenuhi aspek spasial heterogenitas, berlaku juga sebaliknya.

- b. Melakukan pemodelan data prevalensi penyakit kusta di Jawa Timur berdasarkan model GWR dengan tahapan antara lain:

- i. Menentukan *latitude* dan *longitude* di setiap kabupaten dan kota di Jawa Timur.
- ii. Melakukan perhitungan jarak *Euclidean* antar pengamatan berdasarkan lokasi geografis.
- iii. Menghitung nilai *bandwidth* optimum menggunakan kriteria AIC.
- iv. Memilih matriks pembobot dengan menggunakan fungsi *kernel Gaussian*.
- v. Mengestimasi parameter model GWR dengan menggunakan *bandwidth* optimum, yang dijelaskan pada model berikut ini:

$$\hat{Y} = \beta_0(u_i, v_i) + \sum_{k=1}^3 \beta_k(u_i, v_i) x_{ik}$$

- vi. Menguji kesesuaian model GWR.
 - vii. Menguji signifikansi parameter secara parsial.
3. Menganalisis model GWR terkait prevalensi kasus penyakit kusta di setiap kabupaten dan kota di Jawa Timur dengan peta tematik
 - a. Melakukan pemodelan prevalensi penyakit kusta di Jawa Timur dengan pendekatan model regresi linier global dengan tahapan antara lain:
 - i. Mengestimasi model regresi global dengan data prevalensi penyakit kusta di Jawa Timur sebagai variabel respon dengan metode *Ordinary Least Square* (OLS), yang dijelaskan pada model berikut ini :
 12. $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$
 - ii. Menguji asumsi klasik yang berlaku pada model regresi global.
 - iii. Menguji signifikansi parameter secara bersama maupun parsial.

- b. Membandingkan hasil akurasi dan kebaikan model antara model regresi global dan model GWR menggunakan ukuran koefisien determinasi, koefisien determinasi terkoreksi, dan AIC.
- c. Menginterpretasikan model terbaik dan menyimpulkan hasil yang diperoleh dengan peta tematik.

HASIL DAN PEMBAHASAN

Deskripsi Data

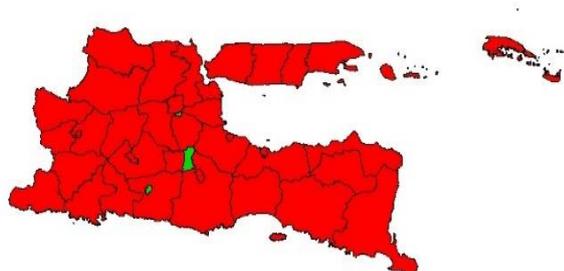
Sebelum melakukan pemodelan, ditampilkan nilai statistika secara deskriptif dari data yang digunakan pada Tabel 1.

Tabel 1. Deskripsi Statistika.

Variabel	Mean	Maks	Min
Y	0,422	2,06	0
X_1	17,928	55,93	1,29
X_2	9,409	23,73	1,65
X_3	116,292	265,56	8,09

Berdasarkan Tabel 1 dapat diperoleh informasi bahwa angka prevalensi kasus penyakit kusta di Jawa Timur memiliki rata – rata sebesar 0,422 dengan nilai maksimum sebesar 2.06 yaitu terdapat di Kabupaten Sumenep dan nilai minimum sebesar 0 yaitu terdapat di Kota Blitar, Kota Batu, dan Kota Mojokerto. Dengan kata lain, angka prevalensi rata – rata sebesar 0,422 menyatakan nilai rasio dari jumlah kasus kusta di Jawa Timur dibagi dengan jumlah seluruh populasi dikali 10.000.

Indikator Pencapaian Target Eliminasi Kusta berdasarkan Peraturan Menteri Kesehatan Republik Indonesia Nomor 11 Tahun 2019 ayat 1 sampai 3 menyatakan bahwa target angka prevalensi penyakit kusta adalah sebesar kurang dari 0,0001. Oleh karena itu, klasifikasi prevalensi penyakit kusta di Jawa Timur dilakukan dengan 2 kriteria yaitu memenuhi target dan tidak memenuhi target sebagaimana digambarkan dalam peta tematik pada Gambar 1.



Gambar 1. Peta Tematik Prevalensi Kasus Penyakit Kusta di Jawa Timur

Pada Gambar 1, daerah yang memenuhi target eliminasi kusta ditandai dengan warna hijau sedangkan daerah yang belum memenuhi target eliminasi kusta ditandai dengan warna merah.

Pemodelan Prevalensi Kasus Penyakit Kusta dengan Metode GWR

Sebelum melakukan pemodelan GWR, asumsi spasial dalam data harus terpenuhi terlebih dahulu. Pengujian asumsi spasial secara lengkap disajikan pada Tabel 2. Kriteria pengambilan kesimpulan yang digunakan adalah asumsi spasial terpenuhi apabila nilai p-value kurang dari tingkat signifikansi 10%.

Tabel 2. Pengujian Asumsi Spasial

Asumsi	Pengujian	p-value	Kesimpulan
Dependensi spasial	<i>Moran's I</i>	5,20e-07	Terdapat dependensi spasial dalam data
Heterogenitas spasial	<i>Breusch Pagan</i>	0,001347	Terdapat heterogenitas spasial dalam data

Berdasarkan Tabel 2 dapat dilihat bahwa semua asumsi spasial telah terpenuhi sehingga analisis spasial dapat dilanjutkan. Setelah dilakukan pengujian asumsi spasial, analisis dapat dilanjutkan dengan pemilihan *bandwidth* optimum. Pemilihan nilai *bandwidth* secara optimum tersebut akan berpengaruh terhadap ketepatan model yang dibentuk. Rentang nilai *bandwidth* diperoleh dari nilai minimum dan maksimum jarak *Euclidean* pada setiap lokasi pengamatan. Dalam hal ini, nilai rentang *bandwidth* tersebut adalah 0,01414214 dan 3,150016.

Penentuan pembobot fungsi *kernel* dilakukan dengan membandingkan nilai AIC yang diperoleh berdasarkan fungsi *kernel* yaitu *Fixed Gaussian*, *Fixed Bisquare*, *Adaptive Gaussian*, dan *Adaptive Bisquare*. Namun, pada kasus ini, fungsi pembobot *Adaptive Gaussian* dan *Adaptive Bisquare* tidak dapat dilakukan karena tidak konvergen pada nilai

rentang *bandwidth* yang didefinisikan. Oleh sebab itu, dipertimbangkan penggunaan dua fungsi pembobot *kernel* yaitu *Fixed Gaussian* dan *Fixed Bisquare* yang hasilnya diringkas pada Tabel 3.

Tabel 3. Perbandingan Nilai AIC pada Tiap Pembobot Fungsi *Kernel*

Pembobot	<i>Bandwidth</i> Optimum	Nilai AIC
<i>Fixed Gaussian</i>	0,134	-73,124
<i>Fixed Bisquare</i>	0,965	2,757

Berdasarkan Tabel 3 diperoleh informasi bahwa nilai AIC minimum diperoleh dengan pembobot fungsi *Kernel Fixed Gaussian*. Hal ini berarti pembobot fungsi tersebut yang dipilih untuk mengestimasi model terbaik dalam penelitian ini.

Selanjutnya, untuk menguji apakah model GWR yang dihasilkan berbeda dengan model regresi global, maka dilakukan uji kesesuaian model GWR. Berikut adalah hipotesis yang digunakan:

H_0 : Tidak ada perbedaan signifikan antara model GWR dengan model regresi global.

H_1 : Terdapat perbedaan signifikan antara model GWR dengan model regresi global.

Kriteria pengujian yang digunakan dengan tingkat kesalahan sebesar 10% adalah H_0 ditolak jika $F > F_{(\alpha;df_1;df_2)}$ atau $F > F_{(0.10;34;30,987)}$. Dengan kata lain, H_0 ditolak jika $F > 1,582$. Hasil pengujian tersebut diringkas pada Tabel 4 sebagai berikut.

Tabel 4. Hasil Uji Kesesuaian Model GWR

Keterangan	SS	DF	MS	F
<i>Global Residuals</i>	5,99	34,00		
<i>GWR Improvement t</i>	5,93	30,99	0,19	
<i>GWR Residuals</i>	0,05	3,01	0,02	10,61

Berdasarkan Tabel 4, nilai F yang diperoleh sebesar 10,61 lebih besar dari daerah kritis yaitu 1,582 sehingga keputusannya adalah tolak H_0 . Kesimpulan yang diperoleh adalah terdapat perbedaan signifikan antara model

GWR dengan model regresi global. Hal ini menunjukkan pengaruh spasial cukup berperan dalam pemodelan kasus ini sehingga menghasilkan model yang berbeda dengan regresi global.

Untuk mengidentifikasi variabel prediktor yang berpengaruh signifikan di setiap lokasi, maka diperlukan pengujian signifikansi parameter secara parsial. Pengujian tersebut menggunakan statistik uji T . Kriteria pengujian yang digunakan dengan nilai α sebesar 10% yaitu tolak H_0 jika $|T_{ki}| > t_{(0.05;3.013)}$ atau $|T_{ki}| > 2,34923$. Berdasarkan hasil analisis pada model GWR di setiap kabupaten dan kota di Jawa Timur, maka dapat dikelompokkan variabel prediktor yang berpengaruh signifikan terhadap variabel respon. Pengelompokkan ini diringkas dalam Tabel 5.

Tabel 5. Pengelompokkan Kabupaten/Kota di Jawa Timur Berdasarkan Variabel Prediktor yang Berpengaruh Signifikan

Kabupaten/Kota	Variabel yang Berpengaruh Signifikan
Sidoarjo, Kota Pasuruan	X_1
Pasuruan, Sampang	X_2
Kediri, Jombang, Kota Kediri, Situbondo	X_3
Tuban, Lamongan, Gresik, Bangkalan, Kota Mojokerto, Kota Surabaya	X_1, X_3
Bondowoso, Pamekasan, Sumenep	X_2, X_3
Banyuwangi, Kota Probolinggo	X_1, X_2, X_3

Dengan demikian, dapat persamaan model GWR dengan variabel prediktor yang memberi pengaruh signifikan pada setiap lokasi pengamatan. Berdasarkan persamaan model GWR yang diperoleh, dapat dilakukan interpretasi model GWR. Dalam hal ini, interpretasi ditampilkan untuk Kabupaten Sumenep dan Kabupaten Pamekasan yang memiliki prevalensi penyakit kusta tertinggi dan tertinggi kedua di Jawa Timur dengan model GWR sebagai berikut.

Persamaan Model Kabupaten Sumenep

$$\hat{y} = -1,296176 + 0,102920X_2 + 0,004532X_3$$

Interpretasi Model :

Setiap kenaikan persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak (X_2) di Kabupaten Sumenep sebesar 1% akan menaikkan prevalensi kasus penyakit kusta di wilayah tersebut sebesar 0,102920% dengan variabel lain dianggap konstan, dan setiap kenaikan jumlah penduduk miskin (X_3) di Kabupaten Sumenep sebesar 1 ribu jiwa akan menaikkan prevalensi kasus penyakit kusta di wilayah tersebut sebesar 0,004532% jika variabel lain dianggap konstan.

Persamaan Model Kabupaten Pamekasan

$$\hat{y} = -0,416359 + 0,069750X_2 + 0,002897X_3$$

Interpretasi Model :

Setiap kenaikan persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak (X_2) di Kabupaten Pamekasan sebesar 1% akan menaikkan prevalensi kasus penyakit kusta di wilayah tersebut sebesar 0,069750% dari kasus sebelumnya dengan asumsi variabel lain konstan, dan setiap kenaikan jumlah penduduk miskin (X_3) di Kabupaten Pamekasan sebesar 1 ribu jiwa akan menaikkan prevalensi kasus penyakit kusta di wilayah tersebut sebesar 0,002897% jika variabel lain konstan.

Perbandingan Ukuran Keباikan Model dan Ukuran Akurasi Model GWR dengan Model Regresi Global

Ukuran kebaikan dan akurasi dari model GWR yang telah diperoleh akan dibandingkan dengan model regresi global. Dengan menerapkan teknik OLS, maka diperoleh persamaan model regresi global. Berikut adalah persamaan model regresi global yang diperoleh:

$$\hat{y} = -0,22 + 0,010 X_1 + 0,026 X_2 + 0,001 X_3$$

Sebelum dianalisis lebih lanjut, dilakukan uji asumsi klasik pada model regresi global yang dirangkum dalam Tabel 6.

Tabel 6. Uji Asumsi Klasik pada Regresi Global

Asumsi	Pengujian	Hasil Pengujian
--------	-----------	-----------------

Normalitas Residual	Kolmogorov-Smirnov	Residual berdistribusi normal
Multi-kolinieritas	Melalui nilai VIF	Tidak ada kasus multikolinieritas
Heteroskedastisitas	Glejser	Terdapat gejala heteroskedastisitas

Berdasarkan Tabel 6, terlihat bahwa model regresi global masih memiliki pelanggaran asumsi klasik yaitu terjadi kasus heteroskedastisitas. Selain itu, nilai ukuran evaluasi model pada model GWR memiliki nilai yang lebih baik dibandingkan model regresi global. Ringkasan ukuran evaluasi model tersebut ditampilkan pada Tabel 7. Berdasarkan Tabel 7, diperoleh bahwa nilai *R-Square*, *Adjusted R-Square*, dan nilai AIC pada model GWR lebih baik dibandingkan model regresi global. Hal ini berarti model GWR memiliki ukuran kebaikan model yang lebih baik.

Tabel 7. Perbandingan Ukuran Evaluasi Model

Model	<i>R-Square</i>	<i>Adjusted R-Square</i>	AIC
Model GWR	99,42%	89,39%	-73,12
Model Regresi Global	36,46%	28,76%	47,61

Selain itu, ukuran akurasi model disajikan seperti pada Tabel 8. Berdasarkan Tabel 8, diperoleh bahwa nilai sMAPE dari model GWR lebih tinggi daripada model regresi global yaitu 12,14% pada model GWR dan 28,94% pada model regresi global. Hal ini berarti model GWR memiliki akurasi yang lebih baik sehingga pembahasan lebih lanjut akan didasarkan pada hasil pemodelan GWR.

Tabel 8. Perbandingan Ukuran Akurasi Model

Model	sMAPE
Model GWR	12,14%
Model Regresi Global	28,94%

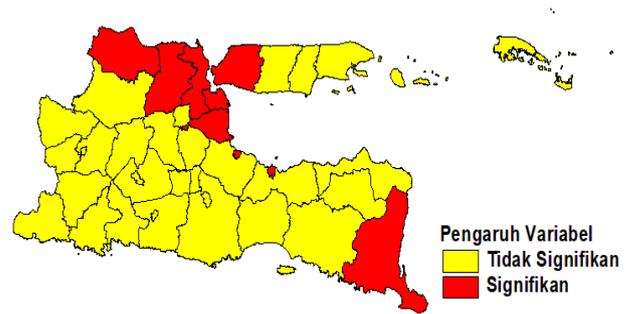
Pembahasan Hasil Pemodelan

Berdasarkan hasil analisis menggunakan GWR, terdapat wilayah kabupaten atau kota di Jawa Timur yang belum memenuhi target eliminasi kusta di Indonesia. Daerah yang

memenuhi target eliminasi kusta berdasarkan nilai dugaan model GWR adalah Kota Blitar. Jika dibandingkan dengan data aktual, kabupaten atau kota yang telah memenuhi target eliminasi kusta adalah Kota Blitar, Kota Mojokerto dan Kota Batu. Hal ini berarti bahwa kabupaten atau kota yang telah memenuhi target eliminasi kusta di Jawa Timur baru mencapai 7.89%. Jika nilai prevalensi penyakit kusta di Jawa Timur dikategorikan berdasarkan tingkat pencapaian target eliminasi kusta yang ditetapkan pemerintah, maka ketepatan klasifikasi yang dihasilkan oleh model GWR adalah sebesar 94.74%.

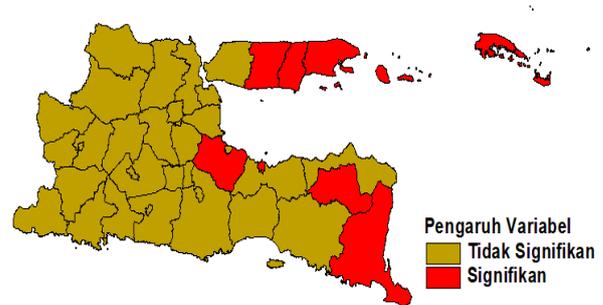
Selain itu, berdasarkan hasil analisis diperoleh bahwa variabel yang digunakan dalam penelitian ini memiliki pengaruh signifikan pada sejumlah wilayah di Jawa Timur. Hasil ini sejalan dengan penelitian Zuhdan dkk. (2017) yang menunjukkan bahwa faktor ekonomi keluarga mempengaruhi jumlah kasus penyakit kusta. Selain itu, hasil penelitian ini mendukung hasil penelitian Muharry (2014) yang menyatakan bahwa kebersihan yang buruk akan mempengaruhi jumlah kasus penyakit kusta. Lebih lanjut, penelitian ini memberikan hasil analisis yang lebih mendalam terkait penyebaran penyakit kusta di Jawa Timur yaitu bahwa tidak seluruh kabupaten dan kota memiliki faktor penyebab terjadinya penyakit kusta yang sama. Dengan kata lain, terdapat wilayah tertentu yang dipengaruhi oleh satu atau beberapa variabel prevalensi penyakit kusta.

Persebaran prevalensi kasus penyakit kusta di Jawa Timur yang secara signifikan dipengaruhi oleh persentase rumah tangga yang menggunakan fasilitas MCK bersama (X_1) adalah sebesar 26,32% yang meliputi Sidoarjo, Kota Pasuruan, Tuban, Lamongan, Gresik, Bangkalan, Kota Mojokerto, Kota Surabaya, Banyuwangi, dan Kota Probolinggo. Persebaran ini disajikan melalui peta tematik pada Gambar 2. Berdasarkan Gambar 2, daerah dengan pengaruh X_1 yang signifikan ditandai dengan warna merah sedangkan daerah dengan pengaruh X_1 yang tidak signifikan ditandai dengan warna kuning.



Gambar 2. Persebaran prevalensi Kasus Penyakit Kusta di Jawa Timur Berdasarkan variabel X_1

Persebaran prevalensi kasus penyakit kusta di Jawa Timur berdasarkan variabel X_2 yaitu persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak disajikan dalam Gambar 3.

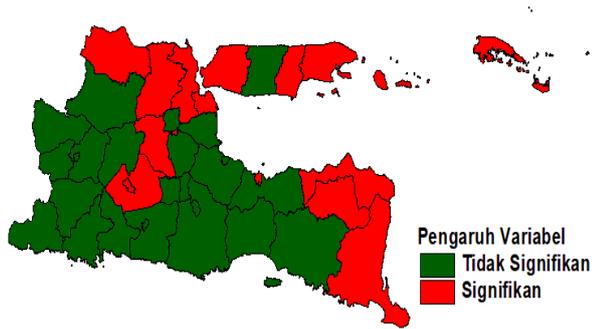


Gambar 3. Persebaran Prevalensi Kasus Penyakit Kusta di Jawa Timur Berdasarkan Variabel X_2

Berdasarkan Gambar 3 dapat diperoleh informasi bahwa 18,42% prevalensi penyakit kusta pada setiap kabupaten dan kota di Jawa Timur secara signifikan dipengaruhi oleh persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak. Daerah tersebut ditandai dengan warna merah yang meliputi wilayah Kabupaten Pasuruan, Sampang, Bondowoso, Pamekasan, Sumenep, Banyuwangi, dan Kota Probolinggo.

Persebaran prevalensi kasus penyakit kusta di Jawa Timur berdasarkan variabel X_3 yaitu jumlah penduduk miskin disajikan pada Gambar 4. Berdasarkan Gambar 4, diperoleh informasi yaitu terdapat 39,47% kabupaten dan kota yang menunjukkan bahwa jumlah penduduk miskin secara signifikan berpengaruh terhadap prevalensi penyakit kusta di Jawa Timur. Daerah ini ditandai dengan warna merah yang meliputi Kediri, Jombang, Kota Kediri, Situbondo, Tuban, Lamongan, Gresik, Bangkalan, Kota Mojokerto, Kota Surabaya, Bondowoso,

Pamekasan, Sumenep, Banyuwangi, dan Kota Probolinggo.



Gambar 4. Persebaran Prevalensi Kasus Penyakit Kusta di Jawa Timur Berdasarkan Variabel X_3

Relevansi Hasil Penelitian dan Rekomendasi Kebijakan

Pada gambaran umum terkait prevalensi penyakit kusta di Jawa Timur, dapat disimpulkan bahwa ada tiga faktor yang berpengaruh yaitu persentase rumah tangga yang menggunakan fasilitas MCK bersama, persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak dan jumlah penduduk miskin. Hal ini mendukung berbagai penelitian terdahulu yang menjelaskan pengaruh ketiga variabel tersebut terhadap prevalensi kasus penyakit kusta yaitu penelitian oleh Zuhdan dkk. (2017) dan Muharry (2014). Dengan demikian, secara umum hasil pemodelan GWR yang diperoleh telah sesuai dan relevan dengan kondisi real.

Hasil pemodelan GWR yang telah diperoleh digunakan sebagai landasan dalam merumuskan rekomendasi kebijakan yang sesuai guna menurunkan angka prevalensi kasus penyakit kusta di Jawa Timur agar sesuai target eliminasi kusta yang ditetapkan oleh pemerintah. Pada bagian ini, rekomendasi kebijakan akan difokuskan pada kabupaten atau kota yang memiliki nilai prevalensi kasus penyakit kusta tertinggi di Jawa Timur yaitu Kabupaten Sumenep. Persamaan model GWR untuk kabupaten Sumenep dengan variabel yang berpengaruh signifikan sebagai berikut:

$$\hat{y} = -1,296176 + 0,102920X_2 + 0,004532X_3$$

Saat ini, persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak (X_2) di Kabupaten Sumenep adalah sebesar 35,26%. Selain itu, jumlah penduduk miskin (X_3) di Kabupaten Sumenep sebesar 17,62 ribu

jiwa dan nilai prevalensi kasus penyakit kusta di Kabupaten Sumenep adalah 2,06. Untuk bisa mencapai target eliminasi kusta dengan prevalensi kasus tertinggi sebesar kurang dari 0,0001, maka persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak (X_2) di Kabupaten Sumenep harus mencapai 11,8191% jika variabel lain dianggap tetap. Di sisi lain, jika jumlah penduduk miskin mencapai nol dengan variabel lain dianggap tetap, maka nilai prevalensi kasus penyakit kusta belum mampu mencapai target yang ditetapkan pemerintah. Hal ini mengindikasikan bahwa diperlukan langkah – langkah yang perlu diterapkan guna mengendalikan variabel yang berpengaruh signifikan tersebut. Secara khusus, Kabupaten Sumenep perlu berfokus pada variabel persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak atau variabel jumlah penduduk miskin. Kombinasi peningkatan melalui kedua variabel tersebut akan menghasilkan penurunan nilai prevalensi penyakit kusta seperti pada Tabel 9.

Tabel 9. Usaha Penurunan Prevalensi Penyakit Kusta di Kabupaten Sumenep

Keterangan	\hat{Y}	X_2	X_3
Data Aktual	2,06	35,26%	17,62
Target Pemerintah	0,0001	12,01%	13,215

Berdasarkan Tabel 9, terlihat bahwa nilai prevalensi penyakit kusta di Kabupaten Sumenep akan mencapai target pemerintah jika nilai persentase rumah tangga yang belum memiliki akses terhadap sanitasi layak (X_2) adalah sebesar 12,01% dan jumlah penduduk miskin (X_3) adalah sebesar 13,215 ribu jiwa. Dengan kata lain, untuk mencapai target eliminasi kusta, Pemerintah Kabupaten Sumenep perlu menurunkan kedua variabel tersebut hingga mencapai nilai tersebut.

Secara umum, rekomendasi kebijakan yang dapat diterapkan oleh pemerintah daerah Jawa Timur adalah memperhatikan variabel yang berpengaruh signifikan tersebut. Dengan memberi fokus perhatian pada variabel yang berpengaruh signifikan di setiap kabupaten dan kota di Jawa Timur, maka prevalensi penyakit kusta di Jawa Timur dapat menurun sehingga

dapat mencapai target eliminasi kusta yang ditetapkan pemerintah.

KESIMPULAN

1. Berdasarkan hasil analisis, diperoleh bahwa prevalensi penyakit kusta tertinggi terjadi di Kabupaten Sumenep dengan nilai 2,06. Kemudian, kabupaten atau kota dengan prevalensi penyakit kusta terendah yaitu terdapat di Kota Blitar, Kota Mojokerto, dan Kota Batu dengan prevalensi penyakit kusta sebesar 0.
2. Model GWR menghasilkan ukuran kebaikan model lebih baik daripada model regresi global yang ditinjau dari koefisien determinasi, koefisien determinasi terkoreksi, dan nilai AIC. Lebih lanjut, ukuran akurasi model GWR juga memberi hasil yang lebih baik dibanding model regresi global yang ditinjau dari nilai SMAPE. Hal ini menunjukkan bahwa faktor spasial berpengaruh terhadap prevalensi penyakit kusta di Jawa Timur. Selain itu, model regresi global juga kurang sesuai untuk diterapkan pada kasus ini karena terjadi penyimpangan asumsi klasik.
3. Hasil pemodelan GWR memberikan persamaan model yang berbeda untuk setiap kabupaten dan kota di Jawa Timur. Persebaran variabel prediktor yang signifikan pada setiap wilayah tersebut adalah variabel X_1 signifikan pada 26,32% kabupaten atau kota, variabel X_2 signifikan pada 18,42% kabupaten atau kota, dan X_3 signifikan pada 39,47% kabupaten atau kota di Jawa Timur.

DAFTAR PUSTAKA

Andreas, C., Harianto, F. Y., Safitri, E. J., dan Chamidah, N. 2021. Analyzing The Effect of BI 7-Days Repo Rate on The Jakarta Composite Index Using Nonparametric Regression Approaches Based on Least Square Spline Estimator. *Jurnal Matematika, Statistika & Komputasi*, 17(3), 447-461.

Badan Pusat Statistik. 2020. *Statistik Indonesia 2020*. Jakarta Pusat: Badan Pusat Statistik.

Badan Pusat Statistik. 2021. *Statistik Indonesia 2021*. Jakarta Pusat: Badan Pusat Statistik.

Bappenas. 2020. *Pedoman Teknis Penyusunan Rencana Aksi: Tujuan Pembangunan Berkelanjutan (TPB)/ Sustainable Development Goals (SDGs). Edisi II*. Jakarta: Bappenas.

Chamidah, N., Saifudin, T., dan Rifada, M. 2014. The vulnerability modeling of dengue hemorrhagic fever disease in surabaya based on spatial logistic regression approach. *Applied Mathematical Sciences*, 8(25), 1369-1379.

Crenata, A. K., Setiawaty, B., dan Ardana, N. K. K. 2012. Pemodelan Nilai Tukar Rupiah Terhadap Dollar Amerika Menggunakan Deret Waktu Hidden Markov Empat Waktu Sebelumnya. *Journal of Mathematics and Its Applications*, 11(2), 37-46.

Diastina, A. R. N., Handajani, S. S., dan Slamet, I. 2019. Analisis Model Geographically Weighted Regression (GWR) pada Kasus Jumlah peserta KB Aktif di Provinsi Jawa Tengah. *Prosiding Seminar Nasional Geotik 2019*. ISSN: 2580-8796. 364-373.

Dziauddin, M. F. dan Idris, Z. 2017. Use Oo Geographically Weighted Regression (GWR) Method to Estimate the Effects of Location Attributes on the Residential Property Values. *The Indonesian Journal of Geography*. 49(1), 97.

Ernawati., Latra, I. Y., dan Puhadi. 2016. Analisis Faktor-Faktor yang Memengaruhi Angka Prevalensi Penyakit Kusta di Jawa Timur dengan Pendekatan *Spatial Durbin Model*. *Jurnal Sains dan Seni ITS*. 5(2): 295-300.

Fatati, I. F., Wijayanto, H., dan Soleh, A. M. 2017. Analisis Regresi Spasial dan Pola Penyebaran Pada Kasus Demam Berdarah Dengue (DBD) di Provinsi Jawa Tengah. *Media Statistika*. 10(2): 95-105.

Indayani, N. I., Windraswara, R., dan Prameswari, G. N. 2017. Analisis Spasial Faktor Risiko Lingkungan dengan Kejadian Kusta di Wilayah Pesisir. *Higeia Journal of Public Health Research and Development*. 1(4),120-130.

Kemenkes. 2019. *Peraturan Menteri Kesehatan Republik Indonesia Nomor 11 Tahun 2019 Tentang Penanggulangan*

- Kusta*. http://hukor.kemkes.go.id/uploads/produk_hukum/PMK_No__11_Th_2019__ttg_Penanggulangan_Kusta.pdf. Diakses pada 18 September 2021.
- Kemenkes. 2021a. *Prevalensi Kusta pada Anak Tinggi. Temukan Kasusnya. Periksa Kontak dan Obati Sampai Tuntas*. <http://p2p.kemkes.go.id/pravalensi-kusta-pada-anak-tinggi-temukan-kasusnya-periksa-kontak-dan-obati-sampai-tuntas/>. diakses pada 18 September 2021.
- Kemenkes. 2021b. *Profil Kesehatan Indonesia 2020*. Jakarta: Kementerian Kesehatan.
- Kominfo Jatim. 2020. *Miliki Penderita Kusta Terbanyak. Masyarakat Diminta Lakukan Deteksi Dini*. <http://kominfo.jatimprov.go.id/read/umum/miliki-penderita-kusta-terbanyak-masyarakat-diminta-lakukan-deteksi-dini>. diakses pada 16 September 2021.
- Kusnandar, D., Debaraja, N. N., dan Fitriani, S. 2021. Pemodelan Sebaran Total Dissolved Solid Menggunakan Metode Mixed Geographically Weighted Regression. *Jurnal Aplikasi Statistika dan Komputasi Statistik*. 13(1), 9-16.
- Lu, B., Charlton, M., Harris, P., dan Fotheringham, S. 2014. Geographically Weighted Regression with a Non-Euclidean Distance Metric: A Case Study Using Hedonic House Price data. *International Journal of Geographical Information Science*. 28(4), 660-681.
- Magri, I. dan Ispriyanti, D. 2017. Pemodelan Data Kemiskinan di Provinsi Sumatera Barat dengan Metode Geographically Weighted Regression (GWR). *Media Statistika*. 6(1), 37-49.
- Muharry, A. 2014. Faktor Risiko Kejadian Kusta. *Jurnal Kesehatan Masyarakat*. 9(2), 174-182.
- Muntasir, M., Salju, E. V., dan Rulianti, L. P. 2018. Studi Faktor-Faktor yang Berhubungan dengan Kejadian Penyakit Kusta pada Wilayah Kerja Puskesmas Bakunase Kota Kupang Tahun 2017. *Jurnal Info Kesehatan*. 16(2), 197-213.
- Olive, D. J. 2017. *Linear Regression*. Gewerbestrasse: Springer.
- Olo, A., Mediani, H. Z., dan Rakhmawati, W. 2021. Hubungan Faktor Air dan Sanitasi dengan Kejadian *Stunting* pada Balita Indonesia. *Jurnal Pendidikan Anak Usia Dini*. 5(2), 1113-1126.
- Paolella, M. S. 2019. *Linear Models and Time-Series Analysis : Regression. ANOVA. ARMA. and GARCH*. New Jersey: John Willey & Sons Inc.
- Permai, S. D., Christina, A., Gunawan, A. A. S. 2021. Fiscal Decentralization Analysis That Affect Economic Performance Using Geographically Weighted Regression (GWR). *Procedia Computer Science*. 179 (2021), 399-406.
- Pertiwi, N. M. S., Sukarsa, I. K. G., dan Susilawati, M. 2020. Pemodelan Jumlah Kasus Penyakit Kusta di Provinsi Jawa Timur. *E-Jurnal Matematika*. 9(1), 42-50.
- Pratiwi, Y. D., Mariani, S., dan Hendikawati, P. 2019. Pemodelan Regresi Spasial Menggunakan Geographically Weighted Regression. *Unnes Journal of Mathematics (UJM)*. 8(2), 32-41.
- Putri, A. S., Pramuningtyas, R., Lestari, N., dan Prakoeswa, F. R. S. 2021. Ketersediaan Jamban Lebih Berpengaruh Dibandingkan Sarana Pembuangan Limbah. Pembuangan Sampah dengan Kusta Wanita di Kabupaten Gresik. *Proceeding Book National Symposium and Workshop Continuing Medical Education XIV*.
- Rasyidah, U. M. 2019. Diare Sebagai Konsekuensi Buruknya Sanitasi Lingkungan. *Jurnal Kesehatan dan Kedokteran*. 1(1), 31-36.
- Ritianty, M. A., Yunus, N., dan Puji. D. 2020. Peran Kelompok Perawatan Diri (KPD) dalam Upaya Mencegah Peningkatan Kecacatan pada Penderita Kusta. *Jurnal Promosi Kesehatan Indonesia*. 15(1), 22-30.
- Sediono, Mardianto, M. F. F., Ulyah, S. M., Pangestu, A. A., Susanti, R., Firdaus, H. A., dan Andreas, C. 2022. The Modelling of Earthquake Magnitude in The Southern Part of Java Island Using Geographically Weighted Regression. *Communications in Mathematical Biology and Neuroscience*. 2022(2022), 13.
- Siswanti dan Wijayanti, Y. 2018. Faktor Risiko Lingkungan Kejadian Kusta. *Higeia Journal of Public Health Research and Development*. 2(3), 252-262.

- Tami, M. 2019. Hubungan Antara Kusta Tipe Pausi Basiler dengan Angka Keberhasilan Pengobatan Kusta di Jawa Timur. *Jurnal Berkala Epidemiologi*. 7(1), 17-24.
- Utami, T. W., Rohman, A., dan Prahutama, A. 2016. Pemodelan Regresi Berganda dan Geographically Weighted Regression pada Tingkat Pengangguran Terbuka di Jawa Tengah. *Media Statistika*. 9(2), 133-147.
- Widayaka, G. W., Mustafid., dan Rahmawati, R. 2016. Pendekatan Mixed Geographically Weighted Regression untuk Pemodelan Pertumbuhan Ekonomi Menurut Kabupaten/Kota di Jawa Tengah. *Jurnal Gaussian*. 5(4), 727-736.
- Wohon, S. C., Hatidja, D., dan Nainggolan, N. 2017. Penentuan Model Regresi Terbaik dengan Menggunakan Metode Stepwise (Studi Kasus: Impor Beras di Sulawesi Utara). *Jurnal Ilmiah Sains*. 17(2), 80-88.
- Wuryandari, T., Hoyyi, A., Kusumawardani, D. S., dan Rahmawati, D. 2014. Identifikasi Autokorelasi Spasial Pada Jumlah Pengangguran di Jawa Tengah Menggunakan Indeks Moran. *Media Statistika*. 7(1), 1-10.
- Zuhdan, E., Kabulrachman, K., dan Hadisaputro, S. 2017. Faktor-faktor yang mempengaruhi kejadian kusta pasca kemoprofilaksis (studi pada kontak penderita kusta di Kabupaten sampang). *Jurnal Epidemiologi Kesehatan Komunitas*. 2(2), 89-98.

PANDEMI COVID-19 DAN *TURNOVER* KE PEKERJAAN INFORMAL: ANALISIS DATA GOOGLE TRENDS

Ari Purwanto Sarwo Prasojo¹

¹Program Studi Statistika, FMIPA Universitas Sebelas Maret, Surakarta
e-mail: ¹rr_ssh@staff.uns.ac.id

Received: January 30, 2022; Revised: April 5, 2022; Accepted: May 25, 2022; Published: June 30, 2022

Abstrak

Pandemi COVID-19 telah berdampak terhadap keberlangsungan pekerjaan bagi pekerja di Indonesia seperti penurunan pendapatan hingga pemutusan hubungan kerja (PHK). Menurunnya permintaan tenaga kerja memungkinkan pekerja terdampak PHK berpindah ke sektor informal. Dengan menggunakan data Google Trends, studi ini bertujuan untuk mengeksplorasi adanya sinyal pergeseran pekerjaan (*job turnover*) ke pekerjaan informal pada periode awal pandemi COVID-19, terutama saat pembatasan fisik dan normal baru di Indonesia. Empat kategori kata kunci: PHK, situs lowongan pekerjaan (loker), kurir, driver online, dan berjualan digunakan sebagai analisis terkait PHK dan pencarian kesempatan kerja yang bersifat informal. Hasil analisis tren dengan menggunakan regresi lokal (LOESS) dan metode difference-in-differences (DD) menemukan adanya sinyal *turnover* saat periode awal pandemi COVID-19, terutama setelah diberlakukannya pembatasan fisik. Sinyal *turnover* tersebut ditunjukkan oleh meningkatnya intensitas penelusuran terkait PHK yang diikuti dengan meningkatnya intensitas penelusuran terkait kesempatan pekerja atau usaha seperti cara berjualan online dan lowongan kurir.

Kata kunci: COVID-19, Google Trends, informal, normal baru, PHK

Abstract

The COVID-19 pandemic has impacted the labor market in Indonesia, such as decreasing wages and layoffs. The decline in worker demand may drive job turnover into the informal sector. Using Google Trends data, this study aims to explore the existence of informal job turnover signals during the early phase of the COVID-19 pandemic, particularly during physical restriction and new normal period in Indonesia. Four keyword categories: "PHK", "situs loker", "kurir", "driver online", and "berjualan" were used to analyze layoff and search the informal job opportunity. The trend analysis results using local regression (LOESS) and difference-in-differences (DD) methods found a signal of informal job turnover in the early phase of the COVID-19 pandemic, especially after physical restriction implementation. The job turnover signal was shown by increasing search intensity about "PHK" (layoff), followed by the search intensity of job opportunities such "berjualan online" (online selling) and "lowongan kurir" (courier jobs).

Keywords: COVID-19, Google Trends, informal, new normal, layoff

PENDAHULUAN

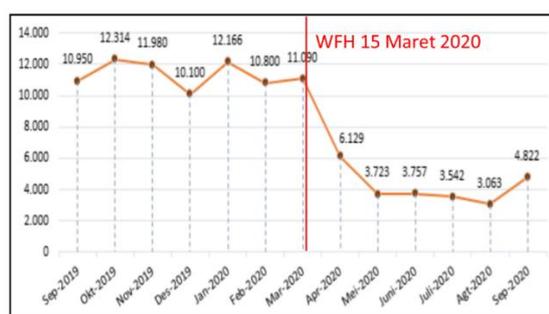
Pandemi COVID-19 telah berlangsung selama dua tahun sejak kasus pertama dikonfirmasi di Indonesia. Tidak hanya masalah kesehatan, pandemi COVID-19 juga telah memberikan dampak yang signifikan terhadap aspek sosial-ekonomi. Pandemi masih menjadi masalah secara global dan tidak sedikit negara yang masih berjuang menghadapinya. Beberapa upaya yang bersifat medis maupun non-medis telah dilakukan oleh negara-negara dalam menekan penyebaran kasus. *Lockdown* merupakan salah satu bentuk intervensi non-medis yang diterapkan oleh negara-negara dalam menekan penyebaran COVID-19.

Meskipun menjadi salah satu upaya yang untuk menekan penyebaran COVID-19, namun *lockdown* memberikan pro dan kontra dalam penerapannya (Greyling et al., 2021). *Lockdown* dapat menekan penyebaran COVID-19 melalui pembatasan mobilitas dan dapat berdampak positif terhadap lingkungan seperti kualitas udara (Mahato et al., 2020). Namun, penerapan *lockdown* juga berdampak negatif terhadap kesejahteraan mental (Brodeur et al., 2021), ekonomi maupun ketenagakerjaan (Vyas, 2020). Tidak semua negara menerapkan *lockdown* sebagai upaya untuk menekan penyebaran COVID-19 karena adanya pro dan kontra tersebut.

Di Indonesia, kebijakan serupa *lockdown* diambil untuk mengurangi laju penyebaran COVID-19. Kebijakan tersebut adalah pemberlakuan *work from home* (WFH) atau pembatasan fisik yang selanjutnya ditingkatkan menjadi pembatasan sosial berskala besar (PSBB). Kebijakan ini disebut serupa *lockdown* karena dalam praktiknya tidak dilakukan pembatasan secara total, melainkan secara parsial. Penerapan kebijakan ini banyak menuai pro dan kontra karena berdampak pada sosial-ekonomi masyarakat. Salah satu dampak yang dirasakan adalah keberlangsungan pekerjaan dan penurunan pendapatan oleh pekerja (Hidayati et al., 2020; Ngadi et al., 2020).

WFH/pembatasan fisik dan PSBB mengakibatkan lesunya aktivitas ekonomi

sehingga menurunkan pendapatan bagi pelaku usaha yang pada ujungnya berdampak pada pemotongan upah, permintaan tenaga kerja hingga pengurangan pekerja (PHK). Jumlah lowongan kerja yang menggambarkan permintaan tenaga kerja mengalami penurunan saat diterapkannya pembatasan fisik (Gambar 3). Hasil survei BPS terhadap pelaku usaha menunjukkan 8,76% perusahaan berhenti beroperasi dan 24,31% beroperasi dengan pengurangan input (jam kerja, mesin, tenaga kerja) (BPS, 2020a). Survei bersama yang dilakukan oleh Pusat Penelitian Kependudukan-LIPI (sekarang Pusat Riset Kependudukan-BRIN), Lembaga Demografi, dan Kemenaker mencatat sebanyak 15,6% pekerja di Indonesia mengalami PHK selama awal penerapan pembatasan sosial, bahkan 13,8% di antaranya tidak mendapatkan pesangon (Ngadi et al., 2020).



Gambar 3. Jumlah Iklan Lowongan Kerja selama Pandemi COVID-19.

Sumber: BPS, 2020b (dimodifikasi).

Keterangan: Jumlah lowongan diperoleh dengan pendekatan pendekatan web scraping pada iklan lowongan pekerjaan.

Turunnya permintaan tenaga kerja dan pengurangan tenaga kerja dapat menimbulkan dampak negatif. Pekerja yang terdampak PHK dimungkinkan mencari pekerjaan alternatif yang lebih mudah untuk menerimanya. Pada kondisi yang terdesak, biasanya pekerja akan bekerja karena keterpaksaan dan sektor informal biasanya menjadi jawaban atas kondisi ini. Hal ini mungkin mengakibatkan terjadinya

pergeseran pekerjaan (*job turnover*) dari sektor formal ke sektor informal.

Berdasarkan pengetahuan penulis, data mengenai *job turnover* pada masa pandemi COVID-19 di Indonesia masih terbatas. Atas hal tersebut, studi ini bertujuan untuk mengeksplorasi sinyal *job turnover* ke sektor informal di masa pandemi COVID-19 dengan menggunakan pendekatan yang unik, yakni menggunakan data Google Trends. Bagian selanjutnya dari artikel ini disusun sebagai berikut. Bagian kedua menjelaskan tentang data dan sampel untuk analisis. Bagian ketiga menguraikan tentang metode analisis data. Hasil dan pembahasan disajikan pada bagian keempat. Terakhir, bagian kelima adalah kesimpulan dan saran.

DATA DAN SAMPEL

Data Google Trends

Google Trends merupakan salah satu sumber big data yang telah banyak dimanfaatkan dalam penelitian sosial (Jun et al., 2018; Mangono et al., 2021; Prasojo, Kusumaningrum, et al., 2020) dan juga dalam studi ketenagakerjaan (Naccarato et al., 2018; Nagao et al., 2019; Simionescu et al., 2020). Pemanfaatan sumber data ini merupakan adaptasi dari perkembangan teknologi digital. Kelebihan data ini adalah tidak memakan waktu dan biaya dan lebih *up to date*. Selain itu mendukung penerapan protokol kesehatan selama pandemi karena tidak perlu berinteraksi dengan subyek penelitian.

Google Trends menyediakan data agregat dari volume penelusuran istilah atau kata kunci tertentu yang dicari melalui mesin pencarian Google. Volume penelusuran yang disajikan berupa volume relatif atau dapat disebut sebagai *Google Relative Search Rate* (GRSR) (Ding et al., 2020). Ukuran ini merupakan indeks yang menggambarkan aktivitas penelusuran relatif untuk suatu istilah atau kata kunci dalam periode waktu dan wilayah geografis tertentu. Unit geografis yang disediakan untuk agregasi volume penelusuran meliputi global, negara, provinsi, metro, dan kota (Rogers, 2016; Stephens-

Davidowitz & Varian, 2014). Namun data deret waktu yang disediakan Google Trends untuk Indonesia adalah tingkat nasional (negara) dan provinsi.

Google Trends menyajikan data harian untuk penelusuran dalam waktu maksimum 3 bulan, data bulanan untuk periode waktu 3 tahun atau selebihnya, dan data mingguan untuk periode waktu lainnya (Stephens-Davidowitz & Varian, 2014). Dalam studi ini, data Google Trends yang dianalisis adalah dalam deret mingguan (lihat penjelasan di bagian tentang sampel). Pandang bahwa s_{tuc} indeks penelusuran relatif kata kunci atau istilah di suatu minggu t tahun u untuk wilayah geografis c . s_{tuc} tidak disediakan oleh Google Trends, dihitung sebagai rasio antara volume kata kunci atau istilah terhadap volume total penelusuran untuk waktu dan geografis yang sama. s_c^{\max} adalah nilai maksimum indeks penelusuran relatif dalam periode yang ditentukan. Google Trends menyajikan indeks penelusuran relatif yang telah dinormalisasi ke dalam rentang nilai 0-100 (GRSR), dihitung sebagai $y_{tuc} = s_{tuc} / s_c^{\max} \times 100$ (Berger et al., 2021).

Sampel

Respons terhadap peristiwa (seperti pandemi) dapat ditangkap melalui perilaku penelusuran di internet seperti data Google Trends (Beytía & Infante, 2020). Perilaku atas respons tersebut dapat dicerminkan oleh pola antar waktu dari tren suatu kata kunci penelusuran. Pada studi ini, beberapa kata kunci ditetapkan menurut kategori seperti yang disajikan pada Tabel 2. Kata kunci yang berkaitan dengan PHK digunakan untuk memproksi tren PHK. Kategori kedua berkaitan dengan pencarian pekerjaan atau penyiapan usaha. Kategori situs loker untuk memproksi aktivitas pencarian lowongan kerja melalui situs lowongan pekerjaan. Kurir, driver online, dan berjualan adalah kategori yang digunakan untuk memproksi aktivitas pencarian kesempatan kerja atau usaha yang bersifat informal. Kategori pekerjaan

tersebut dipilih karena bersinggungan dengan sektor digital yang mana sekitar 8 juta dari 70 juta (11,43%) pekerja informal di Indonesia berada di sektor tersebut dan bergantung pada aplikasi seluler untuk pekerjaannya (Rodriguez, 2020). Selain itu, Kategori tersebut diduga menjadi jenis

pekerjaan dengan permintaan (*demand*) yang lebih baik karena bersamaan dengan tumbuhnya perdagangan secara elektronik (*e-commerce*) (Larasati et al., 2021) dan meningkatnya perilaku berbelanja *online* selama pandemi COVID-19 (Firmandani et al., 2021; Statista, 2021).

Tabel 2. Daftar Kata Kunci Penelusuran untuk Data Google Trends

Kategori	Kata kunci	Keterangan
PHK	"phk buruh"+"phk karyawan"+"phk corona"	Kata kunci yang berkaitan dengan PHK.
Mencari pekerjaan/menyiapkan usaha Situs loker	"jobstreet"	Kata kunci yang berkaitan dengan pencarian lowongan pekerjaan melalui situs penyedia lowongan pekerjaan. Jobstreet dipilih karena menjadi situs penyedia lowongan pekerjaan yang populer atau paling sering dikunjungi di Indonesia (Statista, 2019).
Kurir	"lowongan kurir"+"loker kurir"	Kata kunci yang berkaitan dengan pencarian lowongan pekerjaan sebagai kurir.
Driver online	"daftar gojek"+"daftar grab"	Kata kunci yang berkaitan dengan tata cara mendaftar sebagai mitra <i>platform</i> transportasi <i>online</i> .
Berjualan	"modal jualan"+"cara jualan online"	Kata kunci yang berkaitan dengan mempersiapkan usaha berjualan/secara online.

Catatan: tanda petik (“”) diberikan pada kata kunci penelusuran agar memperoleh hasil kueri penelusuran yang persis dengan frasa tersebut, tanda (+) bermakna “OR” (atau) yang berarti hasilnya berupa kueri penelusuran yang memuat salah satu frasa yang dipisahkan oleh tanda tersebut (Stephens-Davidowitz & Varian, 2014)

Data diambil pada tingkat nasional dalam kurun waktu 1 Januari 2019 sampai dengan 31 Oktober 2020. Periode ini dipilih karena mencakup mulai diterapkannya pembatasan fisik dengan dihimbau untuk bekerja dari rumah (WFH) atau belajar dari rumah pada tanggal 15 Maret 2020 (Kemkominfo, 2020; The Jakarta Post, 2020) dan penerapan “new normal” atau normal baru pada tanggal 5 Juni 2020 (Sutrisno, 2020). Pembatasan mobilitas dilakukan secara bertahap dan ketat sejak periode pembatasan fisik, sedangkan mobilitas mulai dilonggarkan pada normal baru dan aktivitas ekonomi perlahan mulai dijalankan kembali. Fokus terhadap dua periode ini diharapkan dapat menggambarkan tren yang lebih jelas. Meskipun data Google Trends dikumpulkan secara *online*, namun tidak

sensitif terhadap masalah kecukupan sampel karena mencakup pengguna internet, terutama pengguna mesin pencarian google yang cukup besar. *Share* pengguna mesin pencarian google di Indonesia adalah sebesar 98,23% (Statista, 2020). Karena itu, data Google Trends dinilai representatif. Namun, data ini tidak dapat menggambarkan karakteristik yang lebih rinci misalnya karakteristik demografi hanya memberikan informasi secara agregat.

Proses pengambilan data dilakukan masing-masing untuk periode 1 Januari – 31 Oktober baik pada tahun 2019 dan 2020. Periode pada tahun 2019 digunakan sebagai pembandingan kondisi tanpa pandemi. Data yang diperoleh tersebut dalam deret mingguan sebab Google Trends menyediakan data mingguan untuk periode penelusuran antara 9 bulan dan 5

tahun. Karena data tren berasal dari dua tahun atau deret yang berbeda (2019 dan 2020), sehingga faktor skala yang digunakan untuk menghitung GRSR tidak sama. Karena itu, penskalaan ulang diperlukan agar kedua tahun dapat dibandingkan.

Penskalaan

Proses penskalaan mengadopsi studi oleh Brodeur et al. (2021), yakni dengan pembobotan dan normalisasi. Pandang $y_{t,2019}$, $y_{t,2020}$ berturut-turut adalah GRSR kata kunci pada minggu ke- t dari hasil penelusuran antara 1 Januari – 31 Oktober 2019 dan 2020 dengan rata-rata di masing-masing tahun adalah $\overline{y_{.,2019}}$ dan $\overline{y_{.,2020}}$. GRSR kata kunci dari hasil penelusuran antara 1 Januari 2019 – 31 Oktober 2020 adalah $y_{t,2019-2020}$ dan rata-rata GRSR dalam rentang tersebut adalah $\overline{y_{.,2019-2020}}$. Berdasarkan informasi ini, deretterbobot untuk 2019 dan 2020 diperoleh dengan penghitungan:

$$y_{t,2019}^w = y_{t,2019} \times \frac{\overline{y_{.,2019-2020}}}{\overline{y_{.,2019}}} \text{ dan}$$

$$y_{t,2020}^w = y_{t,2020} \times \frac{\overline{y_{.,2019-2020}}}{\overline{y_{.,2020}}}$$

Selanjutnya, normalisasi untuk mendapatkan GRSR yang telah diskala ulang adalah:

$$y_{t,2019}^{norm} = \frac{y_{t,2019}^w}{\max(y_{.,2019}^w, y_{.,2020}^w)} \text{ dan}$$

$$y_{t,2020}^{norm} = \frac{y_{t,2020}^w}{\max(y_{.,2019}^w, y_{.,2020}^w)},$$

dimana $\max(y_{.,2019}^w, y_{.,2020}^w)$ adalah nilai maksimum di antara kedua deret yang telah dibobot.

METODE ANALISIS

Pemulusan Trend dengan Regresi Lokal (LOESS)

Analisis data Google Trends dilakukan secara deskriptif. Tren hasil penelusuran kata-kata kunci diperhalus dengan metode regresi lokal atau *locally estimated*

scatterplot smoothing (LOESS). Pemulusan ini bertujuan untuk memperjelas pola tren data (Cleveland, 1979). LOESS adalah metode pencocokan kurva yang dikembangkan dari metode regresi klasik baik linier maupun non-linier. Metode ini merupakan teknik non-parametrik yang dapat diterapkan untuk kasus hubungan fungsional antara variabel bebas dan variabel terikat yang belum diketahui secara pasti.

Perbedaan mendasar antara regresi klasik dan LOESS adalah pada pendekatan estimasi. Ide dasar pada LOESS adalah mengestimasi regresi polinomial pada setiap himpunan bagian data sehingga menghasilkan kurva yang mulus (Cleveland & Devlin, 1988). Jika $y_t (t = 1, 2, \dots, n)$ adalah GRSR kata kunci pada minggu ke- t (hasil penskalaan ulang), maka spesifikasi model LOESS adalah

$$y_t = g(t) + e_t \tag{1}$$

di mana g adalah fungsi mulus dan e_t adalah residual yang memiliki rata-rata 0 dan varians yang konstan (σ^2). \hat{y}_t adalah penduga dari $g(t)$ yang diperoleh dengan prosedur sebagai berikut (Jacoby, 2000):

Tahap 1 – menentukan parameter pemulusan $\alpha (0 < \alpha < 1)$ dan derajat polinomial $\lambda (\lambda \in 1, 2)$. Parameter α adalah proporsi ukuran pengamatan yang digunakan untuk estimasi regresi lokal. Banyaknya observasi untuk setiap regresi lokal adalah $q = \alpha n$ (diambil nilai integer). Derajat polinomial 1 atau 2 berturut-turut menunjukkan model linier atau kuadratik yang dispesifikasikan. Selanjutnya, mendefinisikan m titik lokasi $s (s = 1, 2, \dots, m)$ dalam rentang nilai t . Titik lokasi s adalah sama dengan t , dan $m = n$. Titik-titik tersebut digunakan sebagai titik estimasi dalam proses pencocokan regresi lokal.

Tahap 2 – untuk setiap titik estimasi s , dihitung bobot titik terdekat. Pandang

$d_{(t)}(s) = |t - s|$ adalah jarak antara titik estimasi dengan titik observasi t yang telah diurutkan dari nilai terkecil. Selanjutnya, jarak dinormalisasi dengan membagi masing-masing jarak dengan jarak ke- q , $d_{(t)}^*(s) = d_{(t)}(s)/d_{(q)}(s)$. Bobot titik terdekat untuk observasi t dihitung dengan

$$w_t(s) = \begin{cases} \left(1 - |d_{(t)}^*(s)|\right)^3, & \text{untuk } d_{(t)}^*(s) < 1 \\ 0, & \text{selainnya} \end{cases}$$

(2)

Tahap 3 – untuk setiap titik estimasi s , dihitung nilai estimasi y atau $g(s)$. Pertama, dilakukan estimasi koefisien model regresi ($\hat{\beta}_{ks}$) yang meminimumkan fungsi

$$\sum_{t=1}^n w_t(s) \left(y_t - \sum_{k=0}^{\lambda} \beta_{ks} t^k \right)^2.$$

(3)

Berikutnya, nilai estimasi y diperoleh dari nilai prediksi model untuk s , yakni

$$g(s) = \sum_{k=0}^{\lambda} \hat{\beta}_{ks} s^k$$

(4)

Tahap 4 (opsional) – Memperbarui nilai estimasi berdasarkan penduga yang bersifat kekar (*robust*). Pembaruan nilai estimasi dilakukan dengan mengulangi tahap 3 dan mengganti bobot $w_t(s)$ menjadi $w_t^*(s) = r_t(s)w_t(s)$. $r_t(s)$ dihitung dari nilai residual yang dibobot dengan fungsi pembobot bisquare, yakni

$$r_t(s) = \begin{cases} \left(1 - e_t^*(s)\right)^2, & \text{untuk } |e_t^*(s)| < 1 \\ 0, & \text{untuk } |e_t^*(s)| \geq 1 \end{cases}$$

(5)

. Di mana $e_t^*(s) = e_t(s)/6\text{median}(|e_t(s)|)$

dan $e_t(s) = y_t - \sum_{k=0}^{\lambda} \hat{\beta}_{ks} t^k$.

Tahap 5 – mengulangi tahap 2, 3, dan 4 (opsional) untuk seluruh nilai s sehingga diperoleh nilai \hat{y}_t untuk $t=1,2,\dots,n$. Melakukan tahap 4 pada proses estimasi

akan diperoleh nilai estimasi yang bersifat *robust*.

Difference-in-Differences (DD)

Metode DD diimplementasikan untuk memperkuat hasil eksplorasi deskriptif dengan LOESS. Meskipun DD termasuk metode kuasi-eksperimen untuk inferensi kausal (Cunningham, 2021), namun saya hanya menggunakannya untuk estimasi hubungan yang bersifat korelasional antara kebijakan WFH, normal baru dengan GRSR. Unsur utama dalam model DD adalah perlakuan (*treatment*) dan kelompok perlakuan (*treatment group*). Pada studi ini, Kebijakan WFH, normal baru menjadi *treatment* yang diminati. Tahun 2020 diperlakukan sebagai *treatment group* karena terjadi pandemi COVID-19 dan terpapar kebijakan tersebut, sedangkan tahun 2019 diperlakukan sebagai *control group* karena bukan tahun pandemi dan tidak terpapar kebijakan tersebut. Karena itu, spesifikasi yang digunakan adalah DD sederhana, yakni

$$y_t = \beta_0 + \beta_1 \text{Tahun}_t + \beta_2 \text{Kebijakan}_t + \beta_3 \text{Tahun}_t \times \text{Kebijakan}_t + \varepsilon_t$$

(6)

Di mana y_t adalah GRSR pada minggu ke- t (hasil penskalaan ulang), Tahun_t adalah dummy tahun, bernilai 1 jika tahun 2020 dan 0 jika tahun 2019, dan Kebijakan_t adalah dummy status diterapkannya kebijakan, bernilai 1 dan 0 untuk periode sesudah dan sebelum diterapkannya kebijakan. Hasil estimasi koefisien β_3 merepresentasikan pengaruh atau hubungan kebijakan terhadap GRSR yang terkait dengan pencarian pekerjaan. Makna hasil estimasi tersebut adalah perbedaan rata-rata GRSR sebelum dan sesudah diterapkannya kebijakan pada tahun 2020 terhadap rata-rata GRSR pada periode yang sama untuk tahun 2019. Uraian tentang perbedaan (*difference*) rata-rata tersebut dijelaskan oleh tabel berikut

Tabel 3. Uraian model DD berdasarkan persamaan 6

Tahun	Kebijakan		Differ ence $= (2) - (1)$
	Sebelum (Kebijakan _t = 0)	Sesudah (Kebijakan _t = 1)	
2019 (Tahun _t = 0) (a)	β_0	$\beta_0 + \beta_2$	β_2
2020 (Tahun _t = 1) (b)	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	β_3
Differ ence (b)-(a)	β_1	$\beta_1 + \beta_3$	β_3

Periode untuk estimasi DD kebijakan WFH adalah hingga sebelum diterapkannya normal baru, sedangkan periode normal baru dimulai setelah diterapkannya WFH. Model pada persamaan (6) diestimasi dengan menggunakan metode kuadrat terkecil (OLS).

Word Clouds

Word clouds diaplikasikan untuk memvisualisasi kueripenelusuran yang berhubungan dengan kata kunci yang ditetapkan. Word clouds memudahkan dalam memberikan pemahaman terhadap kueri yang relevan tersebut dengan cara menyajikan teks dalam bentuk awan yang ukuran dan warnanya didasarkan pada frekuensi teks (Cirillo, 2016). Google Trends mempunyai fitur untuk memunculkan kueri penelusuran yang berhubungan dengan kata kunci yang dicari. Peringkat kueri yang berkaitan disusun berdasarkan skor relatif. Skor 100 bermakna bahwa kueri terkait paling banyak dicari, sedangkan skor 50 bermakna bahwa kueri terkait yang dicari sebanyak setengah kali dari kueri terpopuler, dan demikian seterusnya (Google, 2022). Dengan menggunakan informasi teks dan skor untuk kueri terkait tersebut, visualisasi word clouds disusun untuk memberikan gambaran tentang kueri penelusuran yang memiliki kemiripan makna terhadap kata kunci penelusuran yang ditetapkan. Skor tersebut digunakan sebagai ukuran dan warna huruf pada word clouds.

Implementasi

Seluruh tahapan pengolahan dan analisis data diimplementasikan dengan menggunakan perangkat lunak R v4.1.0 (R Core Team, 2020). Proses pengambilan data Google Trends dilakukan dengan menggunakan library “gtrends” (Massicotte & Eddelbuettel, 2020). Library “ggplot2” (Wickham, 2016) dan “ggwordcloud” (Le Pennec & Slowikowski, 2019) berturut-turut digunakan untuk eksplorasi tren dengan LOESS dan visualisasi word clouds untuk kueriterkait. Estimasi DD diperoleh melalui fungsi “lm” pada library “stats” dan library “sandwich” (Zeileis et al., 2020) untuk estimasi *robust standard error*. Syntax R tersedia dalam suplemen yang disimpan pada repositori <https://github.com/aripurwantos/job-turnover-gtrends>.

HASIL DAN PEMBAHASAN

Analisis Tren

Analisis diawali dengan membandingkan tren penelusuran kata kunci sebelum dan sesudah penerapan WFH/pembatasan fisik, normal baru di tahun 2020 dan pada periode yang sama untuk tahun 2019. Gambar 4a menyajikan tren penelusuran kata kunci terkait PHK. Tren mengalami kenaikan sejak awal periode pengamatan hingga penerapan WFH/pembatasan fisik dan mulai menurun menjelang hingga penerapan normal baru. Sebagaimana yang diharapkan, tren penelusuran terkait dengan PHK mendukung studi terdahulu yang mengemukakan bahwa pandemi COVID-19 telah berdampak negatif terhadap keberlangsungan pekerjaan bagi para pekerja (Ngadi et al., 2020).

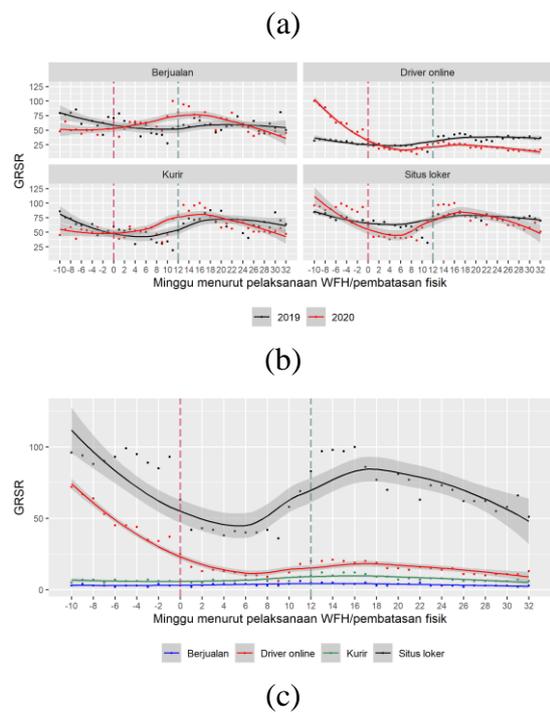
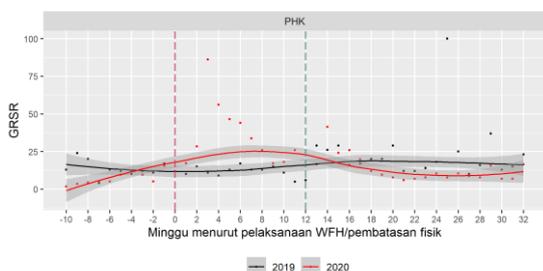
Intensitas penelusuran kata kunci terkait PHK yang mulai meningkat sebelum diterapkannya WFH/pembatasan fisik memberikan indikasi bahwa isu mengenai PHK mulai muncul sejak sebelum pembatasan fisik diumumkan. Wacana pemberlakuan pembatasan fisik telah muncul sebelum pengumuman resmi. Dengan demikian wacana tersebut dimungkinkan menimbulkan beberapa spekulasi atau kekhawatiran

mengenai keberlangsungan sektor usaha atau pekerjaan karena akan diterapkannya pembatasan fisik maupun mobilitas.

Intensitas penelusuran mengenai PHK menurun saat diterapkannya normal baru. Pada periode ini, aktivitas ekonomi mulai diusahakan untuk dipulihkan kembali. Pekerja yang terdampak PHK mungkin berusaha untuk kembali ke pasar kerja. Dugaan ini didukung dengan intensitas pencarian kerja yang digambarkan oleh tren penelusuran terkait situs loker yang meningkat menjelang dan setelah diterapkannya normal baru (Gambar 4b).

Meskipun intensitas penelusuran kerja secara umum mengalami kenaikan hingga penerapan normal baru, namun pencarian kesempatan kerja terkait pekerjaan yang bersifat informal juga mengalami kenaikan. Sebagaimana yang disajikan pada Gambar 4b, tren penelusuran terkait berjualan dan kurir memiliki pola yang serupa. Tren keduanya meningkat setelah diterapkannya WFH/pembatasan fisik dan mencapai volume penelusuran tertinggi saat normal baru.

Tidak seperti penelusuran kesempatan kerja terkait berjualan, kurir, dan situs loker, penelusuran kesempatan kerja terkait driver online cenderung mengalami penurunan hampir di setiap periode (Gambar 4b). Tren mengalami penurunan sejak awal periode pengamatan hingga pertengahan periode WFH/pembatasan fisik. Kenaikan tren terjadi saat normal baru, meskipun volumenya lebih rendah dibandingkan dengan tren di tahun 2019. Namun, penelusuran terkait driver online relatif lebih tinggi jika dibandingkan dengan penelusuran terkait berjualan dan kurir (Gambar 4c).



Gambar 4. (a) Tren Penelusuran Kata Kunci yang Berkaitan dengan PHK; (b) Tren Penelusuran Kata Kunci yang Berkaitan dengan Pencarian Kerja/Mempersiapkan Usaha; (c) Perbandingan Relatif Penelusuran Kata Kunci yang Berkaitan dengan Pencarian Kerja/Mempersiapkan Usaha, tahun 2020.

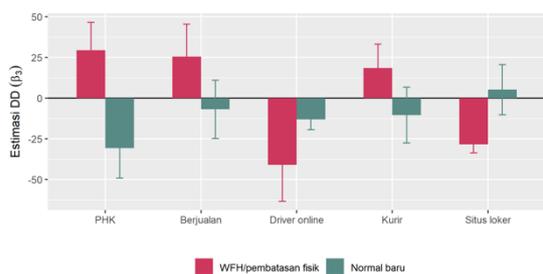
Sumber: Google Trends (diolah penulis).

Keterangan: Kurva mulus mewakili estimasi tren dan 95 persen selang kepercayaan GRSR mingguan. Minggu ke-12 setelah penerapan WFH adalah penerapan normal baru (garis vertikal hijau putus-putus).

Hasil Estimasi DD

Untuk mengukur asosiasi antara kebijakan WFH/pembatasan fisik, normal baru dengan intensitas penelusuran yang berkaitan dengan PHK dan pencarian pekerjaan atau usaha, Gambar 5 menyajikan hasil estimasi Difference-in-Differences (DD). Memperkuat hasil analisis tren, penerapan WFH/pembatasan fisik signifikan meningkatkan intensitas penelusuran terkait PHK dan sebaliknya penerapan normal baru signifikan menurunkan intensitas penelusuran tersebut. Meskipun tidak signifikan, penelusuran terkait situs loker lebih tinggi

saat penerapan normal baru dibandingkan dengan saat WFH/pembatasan fisik. Hal yang sebaliknya terjadi untuk intensitas penelusuran kesem-



Gambar 5. Hasil estimasi pengaruh WFH/pembatasan fisik dan normal baru terhadap penelusuran kata kunci yang berkaitan dengan PHK dan pencarian pekerjaan.

Sumber: Google Trends (diolah penulis).

Keterangan: Bar dan error bar berturut-turut merepresentasikan hasil estimasi DD dan interval kepercayaan 95%.

patan kerja terkait kurir. Intensitas penelusuran mengenai berjualan meningkat dan signifikan saat WFH/pembatasan fisik dan menurun saat normal baru. Pola yang berbeda ditemukan untuk intensitas penelusuran terkait driver online. Penerapan WFH/pembatasan fisik maupun normal baru sama-sama signifikan menurunkan intensi penelusuran tersebut, namun asosiasi yang lebih rendah ditunjukkan saat normal baru.

Kueri Penelusuran Terkait

Untuk mendeskripsikan kueri penelusuran yang terkait dengan kata kunci penelusuran kesempatan kerja dalam studi, Gambar 6 menyajikan daftar kueri terkait dalam bentuk word clouds. "cara jualan online laris" menjadi kueri penelusuran tertinggi yang berkaitan dengan penelusuran terkait berjualan, selanjutnya disusul dengan cara berjualan di *platform marketplace* dan sosial media seperti shopee, lazada, tokopedia, facebook, dan instagram. Munculnya nama kota seperti "jakarta" dan "bandung" pada kueri yang terkait dengan penelusuran kesempatan kerja kurir

menyiratkan bahwa intensitas penelusuran kesempatan kerja ini lebih banyak di kota-kota besar. Hal yang sama ditemukan untuk kueri penelusuran terkait pencarian lowongan pekerjaan melalui situs penyedia lowongan pekerjaan (situs loker).



Gambar 6. Kueri penelusuran yang berkaitan dengan kata kunci dalam studi.

Sumber: Google Trends (diolah penulis).

Diskusi

Keterkaitan antara penelusuran terkait PHK dan pekerjaan informal: sinyal turnover ke perkerjaan informal

Meningkatnya intensitas penelusuran terkait PHK setelah penerapan WFH/pembatasan fisik menyiratkan bahwa gelombang PHK paling besar terjadi saat periode tersebut. Dalam implementasinya, kebijakan WFH/pembatasan fisik merupakan tahap awal menuju pembatasan yang lebih ketat. Kebijakan pembatasan sosial berskala besar (PSBB) yang dikeluarkan melalui Peraturan Pemerintah Nomor 21 Tahun 2020 merupakan salah satu respons signifikan pada periode awal pandemi (Djalante et al., 2020). Kebijakan ini membatasi mobilitas secara lebih ketat, sehingga memengaruhi aktivitas sosial-ekonomi masyarakat. Dampak yang ditimbulkan akibat pembatasan ini adalah menurunnya pendapatan hingga hilangnya pekerjaan karena bekurangnya kegiatan maupun penutupan tempat kerja (Hidayati et al., 2020).

Pada periode yang sama, meningkatnya intensitas penelusuran terkait kesempatan pekerjaan seperti berjualan dan kurir menyiratkan adanya peningkatan intensi terhadap pekerjaan informal pada masa pandemi. Meningkatnya intensi pekerjaan informal dimungkinkan juga karena efek

ikutan setelah terjadinya PHK. Intensi terhadap jenis pekerjaan ini disinyalir dimiliki oleh para pekerja yang terdampak PHK. Karena itu, pola ini dapat menjadi sinyal adanya pergeseran pekerjaan (*job turnover*) menjadi pekerja informal.

Penutupan akses keluar masuk kota besar misalnya Jakarta, termasuk larangan mudik saat diberlakukannya PSBB juga menjadi salah satu hal yang dapat menjelaskan meningkatnya intensitas penelusuran mengenai pekerjaan informal. Studi sebelumnya menemukan bahwa penduduk yang berstatus bekerja baik formal maupun informal adalah kelompok yang paling banyak merencanakan mudik (Prasojo, Aini, et al., 2020). Adapun salah satu motivasinya adalah karena ketidakpastian usaha atau hilangnya pekerjaan karena pandemi (Muhtaruddin & Hartanto, 2020). Ketika akses keluar kota ditutup dan tidak dapat kembali ke kampung halaman, ada potensi bagi kelompok tersebut untuk mencari pekerjaan alternatif sebagai strategi bertahan di kota tempat tinggal.

Teori ekonomi menjelaskan bahwa keputusan bekerja dipengaruhi oleh beberapa faktor. Faktor-faktor tersebut misalnya *reservation wage* dan permintaan dari sektor pekerjaan (Borjas, 2016). *Reservation wage* adalah besaran minimum upah yang dikehendaki oleh pekerja untuk bersedia melakukan pekerjaan tertentu. *Reservation wage* dipengaruhi oleh pendidikan dan upah bukan dari pekerjaan (*non labor income*). Jika individu tidak atau memiliki *non labor income* yang rendah maka *reservation wage* yang dimilikinya cenderung rendah. Demikian halnya untuk individu yang berpendidikan atau berketerampilan rendah yang umumnya memiliki *reservation wage* yang lebih rendah. Selain *reservation wage*, kompensasi terhadap risiko pekerjaan juga menjadi hal yang menentukan seseorang dalam memutuskan untuk bekerja (Borjas, 2016).

Kondisi pandemi beserta kebijakan pembatasan fisik yang menyebabkan turunnya permintaan tenaga kerja serta kondisi kesulitan yang dihadapi pekerja terdampak PHK memberikan implikasi

bekerja karena keterpaksaan. Sebagian besar pekerja dimungkinkan kurang lagi mempertimbangkan *reservation wage* dan kompensasi atas risiko pekerjaan. Pekerjaan dicari dengan pertimbangan untuk penghidupan sementara. Mereka yang terdampak PHK memilih untuk mencoba kembali ke pasar kerja dengan menasar pekerjaan yang relatif mudah dilakukan atau yang mudah menyerap tenaga kerja. Hal inilah yang dapat menjadi salah satu penjelasan meningkatnya intensitas penelusuran yang berkaitan dengan pekerjaan informal tersebut.

Sektor informal disinyalir menjadi tumpuan di masa pandemi, perlu perhatian khusus

Menurunnya intensitas penelusuran terkait driver online karena pembatasan fisik merupakan temuan yang logis. Pembatasan fisik yang disertai dengan pembatasan mobilitas berdampak pada turunnya permintaan jasa transportasi *online*. Meskipun demikian, driver online masih diminati dan dapat menjadi pekerjaan alternatif bagi sebagian yang terdampak PHK. Driver online masih diminati seiring dengan peluang meningkatnya permintaan layanan pembelian makanan secara *online* saat pandemi (Pramisti, 2020) terutama di kota-kota besar. Intensi yang positif untuk penelusuran terkait berjualan dan kurir saat pembatasan fisik menjadi temuan yang relevan. Secara umum, pandemi COVID-19 telah menggeser perilaku belanja dari yang bersifat *offline* menjadi *online*. Hal ini dilakukan untuk mengurangi mobilitas dan intensitas interaksi secara fisik untuk mencegah penularan COVID-19. Karena itu, kesempatan pekerjaan sebagai kurir mungkin meningkat seiring dengan pertumbuhan positif terhadap permintaan layanan jual beli *online* (Safitri, 2020).

Terlepas dari kesempatan yang diberikan pasca PHK maupun saat pandemi oleh alternatif pekerjaan atau usaha tersebut, pekerjaan informal dapat membantu dalam jangka pendek. Pekerjaan tersebut tidak disarankan untuk jangka panjang karena minimnya perlindungan ketenagakerjaan. Sebagai contoh kemitraan antara perusahaan transportasi online (contoh:

gojek, grab vs “driver ojol”) yang seringkali menuai kontroversi. Kemitraan antara keduanya juga seringkali dipandang sebagai kemitraan yang semu, bahkan Frey (2020) menyebut dengan istilah informalisasi yang diformalkan. Perusahaan transportasi *online* seringkali dinilai menerapkan aturan yang menguntungkan perusahaan.

Aturan yang belum sepenuhnya berpihak kepada pekerja (mitra *driver*) dapat menjadikan pekerja semakin rentan di tengah kondisi pandemi. Sulit untuk memastikan kapan pandemi akan berakhir. Jika pandemi tidak berakhir dalam waktu dekat, sektor informal seperti yang telah diulas perlu diberikan perhatian secara khusus untuk mengurangi kerentanan dan meningkatkan kesejahteraan pekerjanya. Hal ini karena sektor-sektor tersebut telah menjadi tumpuan bertahan hidup bagi sebagian besar pekerja yang terdampak PHK, terutama di kota-kota besar.

KESIMPULAN DAN SARAN

Pandemi COVID-19 yang sedang berlangsung memberikan dampak terhadap aspek sosial-ekonomi kehidupan masyarakat termasuk keberlangsungan pekerjaan bagi para pekerja. Analisis terhadap data Google Trends dalam studi ini menemukan sinyal atau indikasi adanya *turnover* pekerjaan ke sektor informal. Adanya PHK membuat pekerja harus beradaptasi dalam menghadapi kesulitan selama masa pandemi. Sinyal adanya *turnover* ke sektor informal ditunjukkan oleh pola intensi penelusuran dengan kata kunci yang terkait dengan beberapa jenis pekerjaan informal pada periode diterapkannya pembatasan fisik. Sinyal ini menyiratkan bahwa sektor informal menjadi salah satu tumpuan bagi pekerja terdampak PHK untuk bertahan selama pandemi. Karena itu, sektor ini perlu lebih diperhatikan untuk mengurangi kerentanan bagi para pekerjanya.

Hasil studi ini perlu diinterpretasikan dengan memperhatikan beberapa keterbatasan. Pertama, intensi masyarakat yang digambarkan oleh pola penelusuran data Google Trends hanya mampu

memberikan gambaran secara agregat. Perlu studi lanjutan untuk melihat pola yang lebih spesifik misalnya menurut karakteristik demografi. Di samping itu, perilaku penelusuran yang digambarkan cenderung merujuk pada masyarakat yang menggunakan internet khususnya mesin pencarian Google. Penduduk umur 19-24 dan 25-49 tahun mendominasi kelompok yang mengakses internet, yakni sekitar 17,13% dan 48,24% (BPS, 2021). Dengan demikian hasil studi ini mungkin bias terhadap kelompok umur tersebut, namun umur tersebut masih mewakili kelompok umur produktif. Kedua, kata kunci yang ditetapkan lebih berkaitan dengan jenis pekerjaan informal yang berhubungan dengan sektor digital dan sebagian besar ditemukan di kota-kota besar. Terakhir, hasil studi ini mewakili penelusuran rata-rata di tingkat nasional dan mungkin menutupi variasi di level geografis yang lebih kecil. Meskipun demikian, studi ini memberikan kontribusi dalam bentuk pengetahuan atau referensi pada studi ketenagakerjaan, terutama yang berkaitan dengan kondisi selama COVID-19 di Indonesia, serta melengkapi data dan informasi yang dikumpulkan melalui survei.

DAFTAR PUSTAKA

- Berger, L. M., Ferrari, G., Leturcq, M., Panico, L., & Solaz, A. (2021). COVID-19 lockdowns and demographically-relevant Google Trends: A cross-national analysis. *PLoS ONE*, 16(3 March), 1–28. <https://doi.org/10.1371/journal.pone.0248072>
- Beytía, P., & Infante, C. C. (2020). *Digital Pathways, Pandemic Trajectories. Using Google Trends to Track Social Responses to COVID-19* (N° 2020-01; HIIG Discussion Paper Series). <http://dx.doi.org/10.2139/ssrn.3571360>
- Borjas, G. J. (2016). *Labor Economics* (7th ed.). Mc Graw Hill.
- BPS. (2020a). *Analisis Hasil Survei Dampak COVID-19 terhadap Pelaku Usaha*. Badan Pusat Statistik.

- BPS. (2020b). *Kajian Big Data sebagai Pelengkap Data dan Informasi Statistik Sosial*. Badan Pusat Statistik.
- BPS. (2021). *Statistik Telekomunikasi Indonesia 2020*. Badan Pusat Statistik.
- Brodeur, A., Clark, A. E., Fleche, S., & Powdthavee, N. (2021). COVID-19, lockdowns and well-being: Evidence from Google Trends. *Journal of Public Economics*, *193*, 104346. <https://doi.org/https://doi.org/10.1016/j.jpubeco.2020.104346>
- Cirillo, A. (2016). *RStudio for R Statistical Computing Cookbook*. Packt Publishing.
- Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836.
- Cleveland, W. S., & Devlin, S. J. (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, *83*(403), 596–610.
- Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.
- Ding, D., del Pozo Cruz, B., Green, M. A., & Bauman, A. E. (2020). Is the COVID-19 lockdown nudging people to be more active: a big data analysis. *British Journal of Sports Medicine*, *54*(20), 1183 LP – 1184. <https://doi.org/10.1136/bjsports-2020-102575>
- Djalante, R., Lassa, J., Setiamarga, D., Sudjatma, A., Indrawan, M., Haryanto, B., Mahfud, C., Sinapoy, M. S., Djalante, S., Rafliana, I., Gunawan, L. A., Surtiari, G. A. K., & Warsilah, H. (2020). Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020. *Progress in Disaster Science*, *6*, 100091. <https://doi.org/https://doi.org/10.1016/j.pdisas.2020.100091>
- Firmandani, W., Sya'bania, A. F., Abdani, F., & Madani, E. (2021). Analysis of Indonesian Consumer Online Shopping Behavior During the Covid-19 Pandemic: A Shopee Case Study. *The International Journal of Business Review (The Jobs Review)*, *4*(2), 191–202.
- Frey, B. (2020). Platform Labor and In/Formality: Organization among Motorcycle Taxi Drivers in Bandung, Indonesia. *Anthropology of Work Review*, *41*(1), 36–49. <https://doi.org/https://doi.org/10.1111/awr.12187>
- Google. (2022). *Menemukan penelusuran terkait*. <https://support.google.com/trends/answer/4355000>
- Greyling, T., Rossouw, S., & Adhikari, T. (2021). The good, the bad and the ugly of lockdowns during Covid-19. *PLOS ONE*, *16*(1), e0245546. <https://doi.org/10.1371/journal.pone.0245546>
- Hidayati, D., Prasajo, A. P. S., Kusumaningrum, D., Surtiari, G. A. K., & Aini, Y. N. (2020). *Riset dampak PSBB: meski menerima bantuan pemerintah, warga hanya mampu bertahan seminggu*. The Conversation. <https://theconversation.com/riset-dampak-psbb-meski-menerima-bantuan-pemerintah-warga-hanya-mampu-bertahan-seminggu-140180>
- Jacoby, W. G. (2000). Loess:: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies*, *19*(4), 577–613. [https://doi.org/https://doi.org/10.1016/S0261-3794\(99\)00028-1](https://doi.org/https://doi.org/10.1016/S0261-3794(99)00028-1)
- Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, *130*, 69–87. <https://doi.org/https://doi.org/10.1016/j.techfore.2017.11.009>
- Kemenkominfo. (2020). *Kominfo Next: Lawan COVID-19. Edisi 15 Maret 2020*. https://eppid.kominfo.go.id/storage/uploads/3_28_3._KNEXT_-_MAR_-

- _2020_-_DRAFT_-_6.pdf
- Larasati, D. N., Bustaman, U., & Pramana, S. (2021). Online Marketplace Data to Figure COVID-19 Impact on Micro and Small Retailers in Indonesia. *Indonesian Journal of Statistics and Its Applications*, 5(2), 333–342. <https://doi.org/10.29244/ijsa.v5i2p333-342>
- Le Pennec, E., & Slowikowski, K. (2019). *ggwordcloud: A Word Cloud Geom for “ggplot2.”* <https://cran.r-project.org/package=ggwordcloud>
- Mahato, S., Pal, S., & Ghosh, K. G. (2020). Effect of lockdown amid COVID-19 pandemic on air quality of the megacity Delhi, India. *Science of The Total Environment*, 730, 139086. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.139086>
- Mangono, T., Smittenaar, P., Caplan, Y., Huang, V. S., Sutermaister, S., Kemp, H., & Sgaier, S. K. (2021). Information-Seeking Patterns During the COVID-19 Pandemic Across the United States: Longitudinal Analysis of Google Trends Data. *J Med Internet Res*, 23(5), e22933. <https://doi.org/10.2196/22933>
- Massicotte, P., & Eddelbuettel, D. (2020). *gtrendsR: Perform and Display Google Trends Queries.* <https://cran.r-project.org/package=gtrendsR>
- Muhtaruddin, & Hartanto, B. D. (2020). Pengaruh Pandemi Covid-19 Terhadap Persepsi Masyarakat Jabodetabek Untuk Melakukan Perjalanan Mudik. *Jurnal Penelitian Transportasi Darat*, 22(2), 107–114. <https://doi.org/10.25104/jptd.v22i2.1657>
- Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. *Technological Forecasting and Social Change*, 130, 114–122. <https://doi.org/https://doi.org/10.1016/j.techfore.2017.11.022>
- Nagao, S., Takeda, F., & Tanaka, R. (2019). Nowcasting of the U.S. unemployment rate using Google Trends. *Finance Research Letters*, 30, 103–109. <https://doi.org/https://doi.org/10.1016/j.frl.2019.04.005>
- Ngadi, Meillianna, R., & Purba, Y. A. (2020). Dampak Pandemi COVID-19 terhadap PHK dan Pendapatan Pekerja di Indonesia. *Jurnal Kependudukan Indonesia, Edisi Khusus*, 43–48.
- Pramisti, N. Q. (2020). *Mitra GoFood Meningkat 40% Selama Pandemi.* Tirto.Id. <https://tirto.id/mitra-gofood-meningkat-40-selama-pandemi-fWhj>
- Prasojo, A. P. S., Aini, Y. N., & Kusumaningrum, D. (2020). Potensi Pola Aliran Mudik pada Masa Pandemi COVID-19. *Jurnal Kependudukan Indonesia, Edisi Spes*, 21–26.
- Prasojo, A. P. S., Kusumaningrum, D., & Dalimunthe, S. A. (2020). Trajektori Pencarian Informasi Warga DKI Jakarta pada Sebelum dan Empat Periode PSBB Menggunakan Google Trends. In *Lesson Learned: Riset Sosial Budaya dalam Pengendalian COVID-19* (pp. 141–174). Yayasan Pustaka Obor Indonesia.
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. <https://www.r-project.org/>
- Rodriguez, D. (2020). *COVID-19 Transforms Informal Work in Indonesia's Thriving Digital Economy.* Asia Pacific Foundation of Canada. <https://www.asiapacific.ca/publication/covid-19-transforms-informal-work-indonesias-thriving>
- Rogers, S. (2016). *What is Google Trends data — and what does it mean?* Google News Lab. <https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8>
- Safitri, K. (2020). *Belanja Online Meningkat, Sektor Logistik Panen Untung.* Kompas.Com. <https://money.kompas.com/read/2020/09/28/112458126/belanja-online-meningkat-sektor-logistik-panen->

- untung
- Simionescu, M., Streimikiene, D., & Strielkowski, W. (2020). What Does Google Trends Tell Us about the Impact of Brexit on the Unemployment Rate in the UK? In *Sustainability* (Vol. 12, Issue 3). <https://doi.org/10.3390/su12031011>
- Statista. (2019). *Most visited job websites across Indonesia as of April 2018*. <https://www.statista.com/statistics/921819/indonesia-most-visited-job-websites/>
- Statista. (2020). *Market share of leading search engines across Indonesia as of August 2020*. <https://www.statista.com/statistics/954420/indonesia-market-share-of-search-engines/>
- Statista. (2021). *Impacts of COVID-19 pandemic on the online purchase behavior among consumers in Indonesia as of May 2020*. <https://www.statista.com/statistics/1127876/indonesia-impact-on-online-purchase-behavior-covid-19/>
- Stephens-Davidowitz, S., & Varian, H. (2014). A hands-on guide to google data. *Further Details on the Construction Can Be Found on the Google Trends Page*.
- Sutrisno, B. (2020). 'Transitional PSBB': A deciding chapter for Jakarta's new normal. *The Jakarta Post*. <https://www.thejakartapost.com/news/2020/06/09/transitional-psbb-a-deciding-chapter-for-jakartas-new-normal.html>
- The Jakarta Post. (2020). *Jokowi calls for 'social distancing' to stem virus spread*. *The Jakarta Post*. <https://www.thejakartapost.com/news/2020/03/15/jokowi-calls-for-social-distancing-to-stem-virus-spread.html>
- Vyas, M. (2020). Impact of Lockdown on Labour in India. *The Indian Journal of Labour Economics*, 63(1), 73–77. <https://doi.org/10.1007/s41027-020-00259-w>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (1st ed.). Springer-Verlag New York. <https://doi.org/10.1007/978-0-387-98141-3>
- Zeileis, A., Köll, S., & Graham, N. (2020). Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R. *Journal of Statistical Software*, 95(1 SE-Articles), 1–36. <https://doi.org/10.18637/jss.v095.i01>

DETERMINAN KELUARNYA PEREMPUAN DARI PASAR TENAGA KERJA PADA MASA PANDEMI COVID-19

Tasmilah Tasmilah¹

¹Politeknik Statistika STIS
e-mail: ¹211709827@stis.ac.id

Received: February 5, 2022; Revised: February 23, 2022; Accepted: May 10, 2022; Published: June 30, 2022

Abstrak

Pandemi Covid-19 mengakibatkan peningkatan jumlah perempuan yang keluar dari pasar tenaga kerja. Dengan menggunakan data sakernas panel tahun 2019-2020, penelitian ini bertujuan untuk mengestimasi faktor penentu perempuan keluar dari pasar tenaga kerja dengan menggunakan regresi logistik. Hasil yang diperoleh bahwa keberadaan balita, status pekerjaan sebagai *self employed*, penggunaan internet, dan tempat bekerja di rumah akan meningkatkan peluang perempuan keluar dari pasar tenaga kerja pada masa pandemi Covid-19. Sedangkan status perempuan sebagai kepala rumah tangga, status pekerjaan sebagai buruh/karyawan, pendidikan tinggi, dan penggunaan teknologi digital akan menurunkan peluang perempuan keluar dari pasar tenaga kerja pada masa pandemi Covid-19.

Kata kunci: Pandemi Covid-19, angkatan kerja perempuan, pengangguran

Abstract

The Covid-19 pandemic has resulted in an increasing number of women leaving the labor market. Using Sakernas data 2019-2020, this study estimates the determinants of women leaving the labor market using logistic regression. The results obtained are that the presence of children under the age of five, employment status as self employed, use of the internet, and a place to work at home will increase women's chances of leaving the labor market during the Covid-19 pandemic. Meanwhile, the status of women as heads of households, employment status as workers/employees, higher education, and the use of digital technology will reduce women's chances of leaving the labor market during the Covid-19 pandemic.

Keywords: Covid-19 pandemi, female labor participant, Unemployment

PENDAHULUAN

Pandemi Covid-19 yang terjadi sejak 2 Maret 2020 mengakibatkan 743.198 orang terinfeksi Covid-19 pada tahun 2020 (Satgas Covid-19, 2021). Untuk mengendalikan penyebaran Covid-19 pemerintah menerapkan pembatasan sosial berskala besar (PSBB) berdasarkan PP No.21 Tahun 2020. Hal tersebut berdampak pada perekonomian nasional sehingga mengakibatkan kontraksi sebesar 2,19 persen pada tahun 2020 (BPS, 2021a). Selain itu, pandemi Covid-19 juga berdampak pada peningkatan jumlah pengangguran sebanyak 2,56 juta orang dan peningkatan Bukan Angkatan Kerja (BAK) sebanyak 0,76 juta orang selama periode Februari-Agustus 2020 (BPS, 2020a). Jumlah pekerja yang keluar dari pasar tenaga kerja karena pandemi Covid-19 paling banyak berjenis kelamin perempuan dengan jumlah mencapai 0,52 juta orang. Pandemi Covid-19 ini terjadi sejak awal Maret 2020 ini mengakibatkan penurunan Tingkat Partisipasi Angkatan Kerja (TPAK) perempuan dari 54,48 pada Februari 2020 menjadi 54,03 pada Februari 2021 (BPS, 2021b). Penurunan TPAK perempuan ini mendorong penulis untuk meneliti faktor-faktor yang menentukan keluarnya perempuan dari pasar tenaga kerja. Menggunakan data ketenagakerjaan sebelum pandemi (2019) dan data ketenagakerjaan pada saat pandemi Covid-19 (2020) akan memotret perubahan status ketenagakerjaan pada perempuan sekaligus mengestimasi karakteristik tenaga kerja yang mempengaruhi perubahan tersebut.

Keluarnya perempuan dari pasar tenaga kerja ini menarik untuk dikaji karena apabila di suatu negara banyak penduduk yang masuk kategori pengangguran atau bukan angkatan kerja maupun keduanya, menunjukkan adanya masalah tenaga kerja yang kurang termanfaatkan (*underutilized*) demikian juga dengan potensi ekonomi suatu negara. Identifikasi *underutilization* tenaga kerja merupakan langkah pertama dalam merancang kebijakan ketenagakerjaan yang bertujuan meningkatkan kesejahteraan penduduk

bekerja, sekaligus mendorong pertumbuhan ekonomi.

Bukan angkatan kerja merupakan salah satu *Key Indicator Labor Market* (KILM) untuk mengukur ketidaktifan secara ekonomi. Salah satu peran KILM dalam analisis tenaga kerja antara lain untuk mempromosikan pekerjaan yang layak (*decent work*) dengan memberikan kesempatan yang sama bagi laki-laki dan perempuan untuk memperoleh pekerjaan produktif dalam kondisi kebebasan, kesetaraan, dan keamanan (ILO, 2016).

Keluarnya perempuan dari pasar tenaga kerja akan semakin meningkatkan ketimpangan gender dalam pekerjaan. Dari perspektif ekonomi, mengurangi kesenjangan gender dalam partisipasi angkatan kerja dapat secara substansial meningkatkan Produk Domestik Bruto (PDB) global dan meningkatkan kesejahteraan individu (ILO, 2017). Hasil penelitian Albanesi (2019) menunjukkan bahwa pertumbuhan berkelanjutan dalam pekerjaan perempuan sejak awal 1990-an berdampak signifikan dalam meningkatkan kinerja ekonomi di Amerika Serikat.

Di era teknologi informasi seperti saat ini pilihan untuk bekerja dari rumah maupun pekerja paruh waktu semakin besar bagi perempuan untuk tetap mengaktualisasikan potensinya dengan bekerja. Aditina & Sugiharti (2019) dengan menggunakan data Sakernas 2018 menemukan bahwa pendidikan tinggi dan pelatihan kerja memberikan peluang yang lebih besar bagi perempuan untuk masuk dalam pasar tenaga kerja. Bahkan, penggunaan teknologi internet memberikan peluang lebih besar bagi perempuan untuk bekerja di rumah. Pada kenyataannya jumlah pekerja paruh waktu perempuan lebih tinggi (36,02 persen) jika dibandingkan dengan pekerja paruh waktu laki-laki (19,39 persen) (BPS, 2020b). Pekerja paruh waktu perempuan mengalami peningkatan dalam tiga tahun terakhir, artinya bahwa menjadi pekerja paruh waktu semakin menjadi pilihan bagi perempuan. Pekerjaan paruh waktu apalagi yang dilakukan di rumah memungkinkan bagi perempuan untuk menyeimbangkan antara

pekerjaan dan urusan rumah tangga. Bahkan bagi perempuan yang memiliki balita cenderung akan memasuki pasar tenaga kerja setelah anaknya cukup aman untuk ditinggal atau dibawa ke tempat kerja (Maryati et al., 2018).

Penelitian terkait kehilangan pekerjaan dan keluarnya perempuan dari pasar tenaga kerja pada masa pandemi Covid-19 telah dilakukan di beberapa negara lain oleh Dang & Nguyen (2021), Alon et al. (2020), Adams-prassl et al. (2020), Kristal & Yaish (2020), dan Hupkau & Petrongolo (2020). Hasil penelitian tersebut mengungkapkan bahwa pandemi Covid-19 memberikan dampak yang besar terhadap kehilangan pekerjaan pada perempuan. Keberadaan balita, status perkawinan, status pekerjaan perempuan, tingkat pendidikan, dan tempat bekerja terbukti berpengaruh terhadap kehilangan pekerjaan pada perempuan pada masa pandemi Covid-19. Adapun penelitian terkait keluarnya tenaga kerja perempuan di Indonesia pada masa pandemi Covid-19 belum ada yang melakukan. Penelitian terhadap determinan keluarnya tenaga kerja perempuan dari pasar tenaga kerja menjadi penting mengingat tingkat partisipasi angkatan kerja perempuan di Indonesia masih rendah. Dengan adanya pandemi Covid-19 ini mengakibatkan TPAK perempuan menjadi semakin rendah sehingga memperlebar ketimpangan gender dalam pekerjaan. Selain itu, menurunnya TPAK perempuan ini harus diantisipasi untuk mencegah penurunan potensi tenaga kerja perempuan terhadap nilai tambah perekonomian.

Berdasarkan diskusi di atas, penelitian ini bertujuan untuk mengkaji determinan keluarnya perempuan dari pasar tenaga kerja pada masa pandemi Covid-19. Dengan menggunakan data panel Sakenas tahun 2019-2020, keluarnya perempuan dari pasar tenaga kerja diamati dari perubahan status bekerja pada tahun 2019 menjadi bukan angkatan kerja pada tahun 2020. Mengetahui variabel yang berpengaruh terhadap keluarnya perempuan dari pasar tenaga kerja diperlukan untuk mencegah kehilangan pekerjaan pada

perempuan yang lebih besar. Terlebih ketika pandemi Covid-19 belum sepenuhnya berakhir, maka pencegahan terhadap kehilangan pekerjaan ataupun keluarnya perempuan dari pasar tenaga kerja bermanfaat untuk meminimalisir penurunan kesejahteraan tenaga kerja perempuan. Mengetahui determinan keluarnya tenaga kerja perempuan pada masa pandemi Covid-19 bisa menjadi masukan bagi pengambil kebijakan agar lebih memperhatikan kondisi tenaga kerja perempuan. Dalam penelitian ini akan dikaji apakah keberadaan balita dan status pekerjaan sebagai *self employed* akan meningkatkan peluang perempuan untuk keluar dari pasar tenaga kerja di tengah pandemi? Demikian juga apakah bekerja di rumah dan penggunaan teknologi digital serta internet akan memperkecil peluang perempuan keluar dari pasar tenaga kerja?

METODOLOGI

1. Tinjauan Referensi

Tenaga kerja merupakan salah satu komponen dari fungsi produksi (selain modal) yang telah dikembangkan oleh Charles Cobb dan Paul Douglas (Nicholson, 2010). Permintaan tenaga kerja ini akan sangat bergantung pada permintaan output produksi, sehingga ketika terjadi penurunan aktivitas perekonomian akan direspon dengan pengurangan tenaga kerja tidak terkecuali pada tenaga kerja perempuan. Keputusan perempuan untuk berpartisipasi dalam pasar tenaga kerja dipengaruhi oleh berbagai faktor. Pendidikan yang meningkat pada perempuan turut meningkatkan partisipasi perempuan dalam pasar tenaga kerja (Borjas, 2012). Selain itu, keterlibatan perempuan dalam pasar tenaga kerja juga didorong oleh motif ekonomi yaitu untuk menambah pendapatan (Dewi, 2012).

Pada masa pandemi Covid-19 meski hampir semua lapangan pekerjaan terkena imbas, namun ada beberapa pekerjaan yang masih bisa dilakukan dari rumah dengan memanfaatkan teknologi informasi. Bahkan selama pembatasan sosial selama pandemi, kegiatan ekonomi maupun pekerjaan lainnya diatur untuk dilakukan di rumah (*work from home*). Dang & Nguyen (2021)

dalam penelitiannya pada masa pandemi di enam negara (China, Jepang, Korea Selatan, Italia, Inggris, dan AS) mengungkapkan bahwa pekerja wanita 24 persen lebih mungkin kehilangan pekerjaan secara permanen daripada pekerja laki-laki. Semakin tua usia perempuan, peluangnya makin kecil mengalami kehilangan pekerjaan. Demikian juga perempuan yang tinggal di perkotaan berpeluang kehilangan pekerjaan dibandingkan di perdesaan.

Alon et al. (2020) mengemukakan bahwa pandemi Covid-19 meningkatkan pengasuhan anak bagi ibu karena penutupan sekolah. Penelitian di negara-negara bagian di Amerika Serikat, dimana pekerja perempuan mengalami lebih banyak kehilangan pekerjaan formal daripada pekerja laki-laki, dan kesenjangan gender dalam pekerjaan semakin melebar dari waktu ke waktu (Alon et al., 2020). Adams-prassl et al. (2020) di AS dan Inggris, perempuan dan pekerja tanpa gelar sarjana secara signifikan lebih mungkin kehilangan pekerjaan mereka. Pekerja yang melakukan pekerjaan dari rumah memiliki kemungkinan yang lebih rendah mengalami kehilangan pekerjaan. Pekerja yang berstatus karyawan/buruh yang memperoleh gaji di Inggris cenderung mengalami kehilangan pekerjaan dibandingkan pekerja yang tidak memperoleh gaji. Namun, karyawan yang memperoleh gaji di AS dan Jerman lebih kecil kemungkinannya untuk kehilangan pekerjaan. *Self employed* di AS dan Inggris cenderung tidak mengalami kehilangan pekerjaan, sedangkan di Jerman diperoleh hasil sebaliknya.

Hasil penelitian dari ILO (2017) perempuan dengan anak-anak di negara berkembang memiliki kecenderungan yang lebih besar untuk aktif dalam pasar tenaga kerja. Perempuan dengan pendidikan menengah di negara berkembang kecenderungannya lebih kecil untuk menjadi angkatan kerja, berbeda dengan perempuan yang berpendidikan tinggi. Sedangkan perempuan yang berstatus kawin kecenderungan untuk menjadi angkatan kerja lebih besar di negara

berkembang, berbeda dengan negara maju dimana kecenderungannya lebih kecil.

Untuk Inggris, Andrew et al. (2020) menemukan bahwa dalam rumah tangga dengan anak-anak yang menjadi tanggungan, ibu lebih mungkin untuk tidak bekerja atau cuti selama krisis daripada ayah. Sevilla & Smith (2020) mendeteksi peningkatan yang lebih besar dalam beban pengasuhan anak secara keseluruhan oleh ibu. Untuk Negara-negara lainnya, Farré et al. (2020) dan Boca et al. (2020) melihat efeknya pada pekerjaan yang dibayar dan pekerjaan rumah tangga di Spanyol dan Italia. Di kedua negara tersebut, perempuan mengambil alih sebagian besar beban pengasuhan anak yang meningkat, akan tetapi pengaruh terhadap pasar tenaga kerja kurang jelas. Penelitian Hupkau & Petrongolo (2020) di Inggris menyatakan bahwa wanita dengan anak usia 0-4 tahun berpeluang lebih besar kehilangan pekerjaan. Demikian juga *Work From Home* (WFH) akan mencegah seorang perempuan kehilangan pekerjaan.

Kikuchi et al. (2021) meneliti dampak Covid-19 terhadap pasar tenaga kerja di Jepang, hasilnya ditemukan bahwa efek negatif dari guncangan COVID-19 di pasar tenaga kerja sangat bervariasi di antara pekerja. Pekerja yang lebih muda mengalami dampak lebih buruk daripada pekerja yang lebih tua, perempuan terdampak lebih buruk daripada laki-laki. Kristal & Yaish (2020) melaporkan bahwa pekerja wanita telah terkena dampak lebih parah daripada pria oleh pandemi Covid-19 di Israel.

Dampak pandemi yang besar pada pekerja perempuan mengakibatkan sangat sedikit perempuan berstatus kawin yang mampu meningkatkan pekerjaan untuk membantu pendapatan dari pasangan yang kehilangan pekerjaan. Dalam jangka pendek dan menengah, resesi akibat pandemi mengikis posisi perempuan di pasar tenaga kerja yaitu melalui kehilangan pekerjaan langsung dan kehilangan pengalaman kerja yang disebabkan rendahnya lapangan pekerjaan selama resesi (Alon et al., 2020).

2. Metode Analisis

Penelitian ini menggunakan analisis regresi logistik multinomial untuk mengestimasi determinan keluarnya perempuan dari pasar tenaga kerja di tengah pandemi. Regresi logistik digunakan jika variabel respon bersifat kategorik (nominal atau ordinal) dengan variabel prediktor bersifat kontinu maupun kategorik. Analisis regresi logistik multinomial merupakan regresi logistik yang digunakan saat variabel terikat mempunyai skala yang bersifat *polichotomus* atau multinomial yaitu berskala nominal dengan lebih dari dua kategori (Hosmer & Lemeshow, 1991). Variabel terikat dalam penelitian ini terdiri atas 3 kategori yaitu tetap bekerja, berubah menjadi bukan angkatan kerja, dan berubah menjadi pengangguran. Bentuk persamaan fungsi logit dalam penelitian ini adalah sebagai berikut:

$$Z_j = \ln\left(\frac{P(Y=j|x)}{P(Y=0|x)}\right) = \beta_{j0} + \beta_{j1} \text{balita} + \beta_{j2} \text{krt} + \beta_{j3} \text{selfemployed} + \beta_{j4} \text{karyawan} + \beta_{j5} \text{dirumah} + \beta_{j6} \text{ed_menengah} + \beta_{j7} \text{ed_tinggi} + \beta_{j8} \text{digital} + \beta_{j9} \text{internet} + \varepsilon$$

Z_j = peluang seorang menjadi j dibandingkan dengan tetap bekerja. Dalam penelitian ini j terdiri dari menjadi pengangguran dan bukan angkatan kerja.

Dalam penelitian ini digunakan analisis efek marginal untuk menyajikan informasi pengaruh atau efek pada masing-masing variabel bebas pada saat variabel lain konstan terhadap peluang tiap kategori pada model logit. Variabel terikat dalam penelitian ini berdasarkan perubahan ketenagakerjaan pada tahun 2019 dan 2020. Perempuan yang tetap bekerja pada tahun 2020 dikategorikan 1 (kategori acuan) sedangkan perempuan yang keluar dari pasar tenaga kerja (menjadi Bukan Angkatan Kerja/BAK) sebagai kategori utama atau kategori 2. Untuk perempuan yang berubah menjadi pengangguran dikategorikan 3.

Variabel bebas dalam penelitian ini adalah keberadaan balita, hubungan dengan kepala rumah tangga (KRT), status bekerja, tempat bekerja, pendidikan, penggunaan teknologi digital, dan penggunaan internet. Definisi operasional dari semua variabel penelitian ini mengikuti konsep dan definisi yang digunakan oleh BPS dalam Survei Angkatan Kerja Nasional (Sakernas)(BPS, 2019).

Penduduk bekerja adalah angkatan kerja yang melakukan pekerjaan dengan maksud memperoleh keuntungan/membantu memperoleh keuntungan selama minimal 1 jam selama seminggu yang lalu secara berturut-turut. Pengangguran adalah angkatan kerja yang tidak bekerja dan sedang mencari pekerjaan atau mempersiapkan usaha atau merasa putus asa memperoleh pekerjaan atau sudah diterima bekerja namun belum memulai. Bukan Angkatan Kerja (BAK) adalah penduduk yang tidak aktif secara ekonomi (tidak bekerja/ sementara tidak bekerja dan bukan merupakan pengangguran). Variabel bebas untuk keberadaan balita apabila perempuan berstatus kawin/cerai dan terdapat balita dalam rumah tangga dan tidak terdapat balita dalam rumah tangga (kategori acuan). Hubungan dengan kepala rumah tangga menjadi perempuan bekerja dengan status sebagai kepala rumah tangga dan bukan sebagai kepala rumah tangga (kategori acuan). Variabel status bekerja dibedakan menjadi berusaha sendiri (*self employed*), karyawan/pegawai/buruh, dan status lainnya (kategori acuan). Untuk tempat bekerja dibedakan menjadi bekerja di rumah dan bekerja di luar rumah (kategori acuan). Tingkat pendidikan dalam penelitian ini didasarkan pada jenjang pendidikan terakhir yang ditamatkan dan memperoleh ijazah/tanda kelulusan yang dibagi menjadi pendidikan rendah (tamat SMP ke bawah) sebagai kategori acuan, pendidikan menengah (tamat SMA/ sederajat), dan pendidikan tinggi. Pada era industri 4.0 maka penggunaan teknologi digital dan internet menjadi suatu kebutuhan sehingga dimasukkan sebagai salah satu variabel utama dalam penelitian ini. Penggunaan teknologi digital

dikategorikan menjadi dua yaitu menggunakan teknologi digital dalam bekerja dan tidak menggunakan teknologi digital dalam bekerja (kategori acuan). Demikian juga dengan penggunaan internet dikategorikan menjadi dua yaitu menggunakan internet dalam pekerjaan dan tidak menggunakan internet dalam pekerjaan (kategori acuan).

Penelitian ini menggunakan data sekunder yang berasal dari Survei Angkatan Kerja Nasional (Sakernas) Agustus 2019 dan Agustus 2020 yang diselenggarakan oleh BPS. Sampel penelitian ini sebanyak 106.542 perempuan pekerja. Perubahan ketenagakerjaan perempuan pada penelitian ini dengan membandingkan status bekerja pada individu yang sama tahun 2019 dan tahun 2020. Variabel bebas yang digunakan merupakan karakteristik tenaga kerja pada tahun 2019, kecuali keikutsertaan program kartu pra kerja berdasarkan kondisi tahun 2020.

Hipotesis dalam penelitian ini adalah keberadaan balita, status sebagai kepala rumah tangga, status pekerjaan, tempat bekerja, tingkat pendidikan, dan penggunaan teknologi digital dan internet berpengaruh signifikan terhadap keluarnya perempuan dari pasar tenaga kerja pada masa pandemi Covid-19. Perempuan dengan balita, berstatus bukan sebagai kepala rumah tangga, berwirausaha, bekerja di rumah, berpendidikan rendah, dan tidak menggunakan teknologi digital/internet akan berpeluang lebih besar keluar dari pasar tenaga kerja pada masa pandemi Covid-19.

HASIL DAN PEMBAHASAN

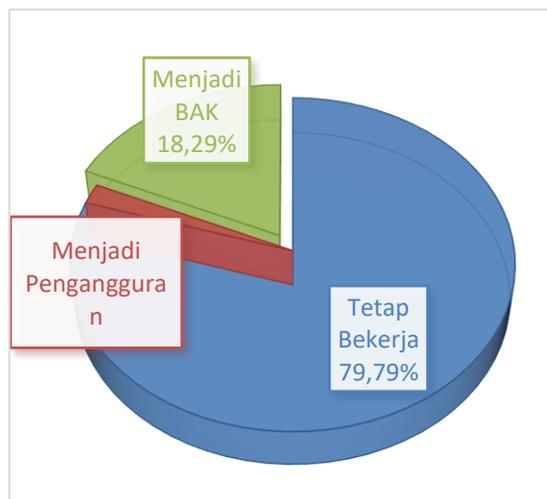
Status pekerjaan utama pada perempuan tahun 2019 disajikan pada gambar 1. Sebagian besar perempuan di Indonesia bekerja sebagai karyawan/buruh/pegawai dengan persentase sebesar 29,73 persen, sebagai pekerja keluarga sebanyak 27,07 persen, dan berstatus sebagai wirausaha mandiri (*self employed*) sebesar 21,88 persen.



Sumber: BPS, 2019

Gambar 1. Persentase Perempuan Bekerja Menurut Status Pekerjaan Utama, Tahun 2019

Pandemi Covid-19 tahun 2020 mengakibatkan resesi ekonomi yang berdampak pada perubahan status ketenagakerjaan pada perempuan sebagaimana ditunjukkan pada gambar 2. Perempuan yang tetap bekerja selama pandemi Covid-19 tahun 2020 sebanyak 79,79 persen, yang berubah menjadi pengangguran sebanyak 1,9 persen, dan keluar dari pasar tenaga kerja/menjadi bukan angkatan kerja sebanyak 18,29 persen.



Sumber: BPS (Data diolah)

Gambar 2. Persentase Tenaga Kerja Perempuan Berdasarkan Status Ketenagakerjaan Tahun 2020

Karakteristik tenaga kerja perempuan pada masa sebelum pandemi

(tahun 2019) dan pada masa pandemi 2020 disajikan pada tabel 1. Pekerja perempuan di Indonesia sebagian besar tidak memiliki balita di rumah. Menurut lokasi kerjanya, hanya 29,64 persen perempuan yang bekerja di rumah dan sisanya 70,36 persen perempuan bekerja di luar rumah. Dalam hal penggunaan teknologi digital, sebagian besar pekerja perempuan tidak menggunakan teknologi digital dan internet. hanya ada 30,35 persen pekerja perempuan yang menggunakan teknologi digital dan 20,04 persen yang menggunakan teknologi internet dalam bekerja. Hal ini bisa jadi karena sebagian besar perempuan di Indonesia berpendidikan rendah, bekerja pada sektor pertanian, dan bertempat tinggal di daerah perdesaan. Pekerja perempuan yang berpendidikan SMP/ sederajat ke bawah mencapai 62,85 persen, yang berpendidikan menengah sebanyak 22,40 persen, dan berpendidikan tinggi hanya 14,75 persen. Adapun menurut lapangan usahanya, pekerja perempuan lebih banyak bekerja pada sektor pertanian

(47,78 persen). Pekerja perempuan yang bekerja di sektor manufaktur sebanyak 30,89 persen dan di sektor jasa sebanyak 21,33 persen.

Pada masa pandemi Covid-19 tahun 2020, tenaga kerja perempuan yang memiliki balita lebih banyak yang keluar dari pasar tenaga kerja (menjadi bukan angkatan kerja) dibandingkan dengan menjadi pengangguran. Sedangkan perempuan yang berubah menjadi pengangguran di tengah pandemi sebagian besar berstatus sebagai karyawan. Hal ini berkebalikan dengan pilihan menjadi bukan angkatan kerja, dimana perempuan yang berstatus sebagai karyawan persentasenya paling sedikit yang keluar dari pasar tenaga kerja.

Perempuan yang berubah menjadi bukan angkatan kerja pada masa pandemi sebagian besar berstatus sebagai pekerja lainnya yaitu dengan rincian pekerja keluarga (31,18 persen), pekerja bebas (6,43 persen), dan wirausaha mandiri (15,69 persen).

Tabel 1: Karakteristik Tenaga Kerja Perempuan Pada Masa Pandemi Covid-19

Karakteristik Tenaga Kerja		Tahun 2019	Tahun 2020		
			Tetap Bekerja	Menjadi BAK	Menjadi Pengangguran
Keberadaan balita	Ya	21,56	23,44	15,4	23,44
	Tidak*	78,44	76,56	84,6	76,56
Hub dengan KRT	KRT	14,56	12,69	10,53	12,69
	Bukan KRT*	85,44	87,31	89,47	87,31
Status bekerja	Wirausaha mandiri	21,89	25,36	18,03	25,36
	Karyawan/pegawai	29,73	19,36	51,71	19,36
	Lainnya*	48,38	55,28	30,26	55,28
Tempat kerja	Di luar rumah*	70,36	66,06	76,27	66,06
	Di rumah	29,64	33,94	23,73	33,94
Pendidikan	Rendah*	62,85	71,60	43,86	71,60
	Menengah	22,40	22,48	41,62	22,48
	Tinggi	14,75	5,91	14,52	5,91
Teknologi Digital	Ya	30,35	20,36	41,72	20,36
	Tidak*	69,65	79,64	58,28	79,64
Penggunaan Internet	Ya	20,04	12,65	29,48	12,65
	Tidak*	79,96	87,35	70,52	87,35

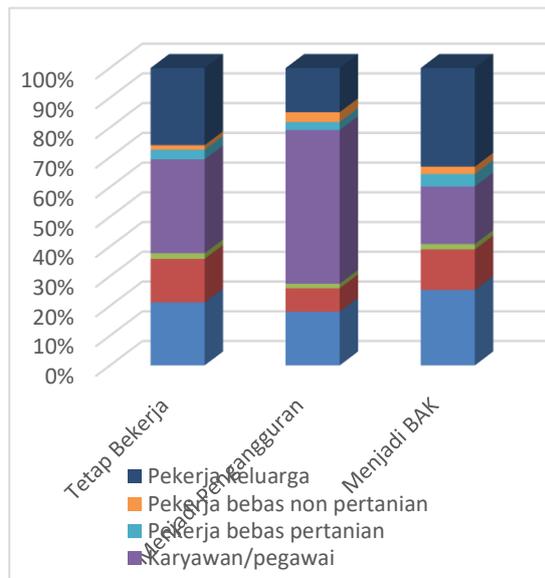
Sumber: Sakernas 2019-2020, BPS (diolah)

Keterangan: *) kategori acuan

Dalam hal penggunaan internet, perempuan yang berubah menjadi bukan angkatan kerja persentase penggunaan teknologi digital dan internetnya paling kecil dibandingkan dengan perempuan yang tetap bekerja maupun berubah menjadi pengangguran.

Sumber: Sakernas 2019-2020, BPS (diolah)
Gambar 3. Perubahan Ketenagakerjaan Perempuan Menurut Status Pekerjaan Utama Tahun 2020

Perubahan ketenagakerjaan perempuan menurut status pekerjaan utama disajikan pada gambar 3. Pekerja perempuan yang tetap bekerja di tengah pandemi pada tahun 2020 sebagian besar berstatus sebagai karyawan/pegawai atau buruh. Untuk perempuan yang berubah menjadi pengangguran pada tahun 2020 didominasi oleh perempuan dengan status sebagai karyawan/pegawai/buruh. Adapun perempuan yang keluar dari pasar tenaga



kerja (bukan angkatan kerja) sebelumnya bekerja sebagai pekerja keluarga/buruh tidak dibayar dan sebagai wirausaha mandiri (*self employed*).

Tabel 2. Hasil Regresi Logistik Multinomial Terhadap Keluarnya Perempuan dari Pasar Tenaga Kerja

Variabel	Menjadi Bukan Angkatan Kerja		Menjadi Pengangguran	
	<i>coeff</i>	<i>P-value</i>	<i>coeff</i>	<i>P-value</i>
Keberadaan balita	0,1411668	0,000***	-0,4099525	0,000***
Status KRT	-0,2996071	0,000***	-0,3478532	0,000***
Status Bekerja				
Wirausaha mandiri	0,0842182	0,000***	0,2760308	0,000***
Karyawan/pegawai	-0,2427391	0,000***	0,9897877	0,000***
Bekerja di rumah	0,0643249	0,001***	0,0747909	0,230
Pendidikan				
Menengah	-0,0098492	0,642	0,6620844	0,000***
Tinggi	-0,8520604	0,000***	-0,4426838	0,000***
Teknologi Digital	-0,3702741	0,000***	-0,0437406	0,554
Penggunaan Internet	0,1153525	0,002**	0,0851328	0,279
Pseudo R ²	: 0,02761			
LR Chi ² (18)	: 3333,48			
Prob Chi ²	: 0,0000			
Jumlah Observasi	: 106.542			

Sumber: hasil olah Stata

Keterangan: *** signifikan pada taraf alpha 1 persen

** signifikan pada taraf alpha 5 persen

Berdasarkan analisis regresi logistik multinomial yang disajikan pada tabel 2, uji serentak dengan statistik uji G dari hasil olah menggunakan STATA 14, diperoleh Likelihood Ratio χ^2 (18) sebesar 3333,48 dan nilai p -value=0,000. Karena nilai *Pearson* yang diperoleh lebih besar dari nilai $\chi^2_{tabel} = 28.86930$ dan p -value lebih kecil dari α 0,05 maka dengan demikian keputusannya adalah tolak H_0 . Hal ini berarti bahwa dengan tingkat kepercayaan 95 persen terdapat pengaruh variabel bebas terhadap variabel terikat, yaitu minimal satu variabel bebas yang berpengaruh terhadap keluarnya tenaga kerja perempuan pada masa pandemi Covid-19. Berdasarkan uji parsial dengan tingkat kepercayaan 95 persen, variabel yang berpengaruh signifikan terhadap keluarnya perempuan dari pasar tenaga kerja (menjadi bukan angkatan kerja) adalah keberadaan balita, status sebagai kepala rumah tangga, status pekerjaan, tempat bekerja, pendidikan tinggi, dan penggunaan teknologi digital dan internet

Berdasarkan analisis efek marginal pada tabel 3, variabel yang berpengaruh paling besar terhadap keluarnya perempuan

dari pasar tenaga kerja adalah keberadaan balita dalam rumah tangga. Sedangkan variabel yang berpengaruh paling besar dalam menurunkan peluang tenaga kerja perempuan menjadi bukan angkatan kerja adalah pendidikan tinggi. Artinya bahwa tenaga kerja perempuan yang memiliki balita di rumah akan berpeluang paling besar menjadi bukan angkatan kerja. Sedangkan tenaga kerja perempuan yang berpendidikan tinggi memiliki peluang paling kecil keluar dari pasar tenaga kerja pada masa pandemi Covid-19.

Adapun variabel yang berpengaruh signifikan terhadap peluang tenaga kerja perempuan menjadi pengangguran adalah keberadaan balita, status sebagai kepala rumah tangga, jenis pekerjaan, dan tingkat pendidikan. Keberadaan balita dalam rumah tangga berpengaruh positif terhadap keluarnya tenaga kerja perempuan dari pasar tenaga kerja pada masa pandemi dan berpengaruh negatif terhadap peluang menjadi pengangguran. keberadaan balita akan meningkatkan peluang perempuan menjadi bukan angkatan kerja sebesar 2,22 persen dan menurunkan peluang tenaga kerja menjadi pengangguran sebesar 0,81 persen.

Tabel 3. Hasil Estimasi Efek Marjinal

Karakteristik Tenaga Kerja		Tetap Bekerja	Menjadi Bukan Angkatan Kerja	Menjadi Pengangguran
Keberadaan balita	Ya	-0,0138	0,0222***	-0,0081***
Hub KRT	KRT	0,0482	-0,0427***	-0,0055***
Status bekerja	Wirausaha mandiri	-0,0158	0,0122***	0,0035***
	Karyawan/pegawai	0,0161	-0,0373***	0,0212***
Tempat bekerja	Di rumah	-0,0103	0,0092***	0,0012
Pendidikan	Menengah	-0,0107	-0,0043	0,0150***
	Tinggi	0,1052	-0,1006***	-0,0046***
Teknologi Digital	Ya	0,0536	-0,0540***	-0,0004
Penggunaan Internet	Ya	-0,0178	0,0166**	0,0012
Pseudo R ²	: 0,02761			
Prob Chi ²	: 0,0000			
Jumlah Observasi	: 106.542			

Sumber: hasil olah Stata

Keterangan: *** signifikan pada taraf alpha 1 persen
** signifikan pada taraf alpha 5 persen

Berbeda dengan perempuan yang tidak memiliki balita, mereka akan aktif mencari pekerjaan lagi atau menyiapkan usaha ketika kehilangan pekerjaan. Pekerja yang memiliki balita di rumah sebagian besar bekerja sebagai karyawan/pegawai/buruh (30,20 persen) dan bekerja sebagai pekerja keluarga/tidak dibayar (28,04 persen). Hasil penelitian ini sama dengan Andrew et al. (2020) di Inggris bahwa perempuan yang memiliki anak akan berpeluang lebih besar untuk tidak bekerja atau mengambil cuti selama pandemi. Demikian juga menurut Hupkau & Petrongolo (2020), perempuan di Inggris yang memiliki anak umur 0-4 tahun lebih berpeluang kehilangan pekerjaan selama pandemi Covid-19. Perempuan yang berstatus sebagai kepala rumah tangga akan menurunkan peluang perempuan keluar dari pasar tenaga kerja ataupun menjadi pengangguran. Perempuan yang berstatus sebagai kepala rumah tangga lebih berpeluang untuk tetap bekerja dibandingkan dengan perempuan yang bukan sebagai kepala rumah tangga. Tanggung jawab nafkah oleh perempuan sebagai kepala rumah tangga lebih besar sehingga lebih memilih untuk tetap bekerja meski harus mengalami transisi dari karyawan/berusaha dibantu buruh menjadi wirausaha sendiri.

Status sebagai wirausaha mandiri berpengaruh positif terhadap keluarnya perempuan dari pasar tenaga kerja (menjadi bukan angkatan kerja) dan menjadi pengangguran. Perempuan yang berstatus sebagai wirausaha mandiri (*self employed*) terdampak paling besar selama pandemi dan berpeluang lebih besar mengalami kehilangan pekerjaan dibandingkan dengan status lainnya. Hal ini sejalan dengan hasil penelitian Adams-prassl et al. (2020) di Jerman, namun berbeda dengan hasil penelitian di Inggris dan Amerika Serikat. Wirausaha perempuan lebih berpeluang menjadi pengangguran atau keluar dari pasar tenaga kerja. Peluang wirausaha

mandiri menjadi bukan angkatan kerja lebih besar dibanding menjadi pengangguran. Perempuan yang berstatus sebagai wirausaha mandiri akan meningkatkan peluang menjadi bukan angkatan kerja sebesar 1,22 persen dan meningkatkan peluang menjadi pengangguran sebesar 0,35 persen. Berbeda dengan perempuan yang berstatus karyawan/buruh/pegawai, ketika kehilangan pekerjaan perempuan ini lebih berpeluang menjadi pengangguran daripada keluar dari pasar tenaga kerja. Perempuan yang berstatus wirausaha mandiri sebagian besar berpendidikan rendah (73,21 persen) dan paling banyak berusaha di sektor manufaktur (54,26 persen). Sedangkan perempuan yang berstatus karyawan/buruh sebagian besar berpendidikan tinggi dan bekerja pada sektor jasa (61,63 persen).

Menurut tempat kerjanya, perempuan yang bekerja di rumah berpeluang lebih besar untuk keluar dari pasar tenaga kerja (menjadi bukan angkatan kerja) selama pandemi Covid-19. Perempuan ini tidak mencari pekerjaan dan tidak menyiapkan usaha baru selama pandemi. Perempuan yang bekerja di rumah lebih berpeluang untuk kehilangan pekerjaan dibandingkan dengan perempuan yang bekerja di luar rumah. Pada kenyataannya perempuan yang bekerja di rumah sebagian besar berpendidikan rendah (68,59 persen), bekerja sebagai wirausaha mandiri (41,08 persen) dan pekerja keluarga (26,36 persen). Menurut lapangan usahanya, perempuan yang bekerja di rumah sebagian besar bekerja pada sektor manufaktur (54,72 persen). Dalam bekerja, perempuan yang bekerja di rumah ini sebagian besar tidak menggunakan teknologi digital (71,70 persen) dan tidak menggunakan internet (85,59 persen).

Menurut tingkat pendidikan, perempuan yang berpendidikan tinggi memiliki peluang yang lebih kecil menjadi bukan angkatan kerja maupun menjadi pengangguran. Pendidikan tinggi yang dimiliki oleh perempuan akan menurunkan peluang keluar dari pasar tenaga kerja

sebesar 10,06 persen dan menurunkan peluang menjadi pengangguran sebesar 0,46 persen. Hasil temuan ini sejalan dengan Adams-prassl et al. (2020). Adapun perempuan dengan pendidikan menengah berpeluang lebih besar untuk menjadi pengangguran. Perempuan yang berpendidikan tinggi sebagian besar bekerja sebagai karyawan/pegawai (84,34 persen), sedangkan perempuan yang berpendidikan rendah sebagian besar bekerja sebagai pekerja keluarga/tidak dibayar (34,58 persen). Hal ini menunjukkan bahwa diperlukan pendidikan yang tinggi pada perempuan agar mampu bertahan di tengah guncangan ekonomi pada masa pandemi. Semakin tinggi pendidikan, maka peluangnya untuk tetap bekerja semakin besar.

Menilik rata-rata lama sekolah perempuan di Indonesia pada tahun 2020 sebesar 8,07 tahun. Artinya bahwa perempuan berusia 25 tahun ke atas di Indonesia rata-rata menempuh pendidikan hingga kelas tiga SMP/ sederajat atau masih berpendidikan rendah. Oleh karena itu perlu mendorong partisipasi perempuan untuk melanjutkan sekolah hingga perguruan tinggi. Untuk jenjang pendidikan menengah (SMA/ sederajat) angka partisipasi sekolah perempuan sebesar 84,92 persen, sedangkan untuk jenjang perguruan tinggi angka partisipasi kasarnya hanya sebesar 32,21 persen pada tahun 2020.

Penggunaan teknologi digital akan menurunkan tenaga kerja perempuan keluar dari pasar tenaga kerja (menjadi bukan angkatan kerja) sebesar 5,40 persen. Hal ini berarti bahwa penggunaan teknologi digital akan menjadikan perempuan lebih tahan untuk tetap bekerja pada masa pandemi dibandingkan dengan perempuan yang tidak menggunakan teknologi digital dalam bekerja. Adapun penggunaan internet dalam bekerja berpengaruh positif terhadap keluarnya perempuan dari pasar tenaga kerja dibandingkan dengan perempuan yang tidak menggunakan internet. Penggunaan internet akan meningkatkan peluang perempuan keluar dari pasar tenaga kerja sebesar 1,66 persen. Pada kenyataannya, perempuan yang bekerja

menggunakan internet ini sebagian besar berpendidikan tinggi dan bekerja sebagai karyawan di sektor jasa.

KESIMPULAN

Pandemi Covid-19 yang terjadi pada tahun 2020 mengakibatkan penurunan perekonomian hingga 2,79 persen dan membuat banyak pekerja kehilangan pekerjaan tidak terkecuali pekerja perempuan. Ada sebanyak 1,93 persen pekerja perempuan yang statusnya berubah menjadi pengangguran dan 18,29 persen pekerja perempuan yang keluar dari pasar tenaga kerja (bukan Angkatan kerja). Dari hasil analisis regresi logistik multinomial, dapat disimpulkan bahwa keberadaan balita dalam rumah tangga berpengaruh paling besar terhadap keluarnya perempuan dari pasar tenaga kerja di tengah pandemi. Perempuan dengan status wirausaha mandiri berpeluang lebih besar mengalami kehilangan pekerjaan pada masa pandemi dan menjadi bukan angkatan kerja. Adapun perempuan yang berstatus karyawan/pegawai/buruh lebih berpeluang untuk mencari pekerjaan yang baru (menjadi pengangguran) ketika kehilangan pekerjaan dibandingkan menjadi bukan angkatan kerja. Menurut tempat bekerjanya, perempuan yang bekerja di rumah lebih berpeluang untuk menjadi bukan angkatan kerja, hal ini karena sebagian besar perempuan yang bekerja di rumah berstatus sebagai wirausaha mandiri, berpendidikan rendah, dan bekerja pada sektor pertanian. Menurut tingkat pendidikan, semakin tinggi pendidikan maka semakin kecil peluangnya mengalami kehilangan pekerjaan selama pandemi. Demikian juga, semakin tinggi pendidikan semakin kecil peluang untuk keluar dari pasar tenaga kerja (menjadi bukan angkatan kerja).

Meningkatkan pendidikan bagi perempuan perlu dilakukan agar mampu bertahan dalam pekerjaan di tengah guncangan ekonomi akibat pandemi. Selain itu, penggunaan teknologi digital akan menurunkan peluang tenaga kerja perempuan keluar dari pasar tenaga kerja. Memberikan pelatihan digital pada tenaga kerja perempuan terutama pada wirausaha mandiri akan meningkatkan peluang bagi

perempuan untuk tetap bekerja di tengah pandemi. Penggunaan teknologi digital ini memungkinkan perempuan untuk tetap bekerja meskipun memiliki balita dengan tempat bekerja di rumah. Mengingat dalam penelitian ini variabel penggunaan internet dalam bekerja ditemukan meningkatkan peluang seorang pekerja perempuan keluar dari pasar tenaga kerja di tengah pandemi Covid-19, maka untuk penelitian selanjutnya perlu dibedakan tujuan penggunaan internet apakah untuk komunikasi, pemasaran, atau untuk penjualan. Hal ini penting agar nantinya diperoleh kesimpulan yang lebih tepat dan akurat terkait pengaruh penggunaan internet dalam bekerja selama pandemi Covid-19.

REFERENSI

- Adams-prassl, A., Boneva, T., Golin, M., & Rauh, C. (2020). Inequality in the Impact of the Coronavirus Shock: Evidence from Real Time Surveys. *IZA – Institute of Labor Economics*, 13183, 51.
- Aditina, N., & Sugiharti, L. (2019). Pengaruh Teknologi Internet Terhadap Keputusan Perempuan untuk Bekerja di Rumah. *Jurnal Economia*, 15(2), 209–220.
- Albanesi, S. (2019). *Changing Business Cycles: The Role Of Women's Employment* (25655, Vol. 52, Issue 1).
- Alon, T., Doepke, M., Olmstead-Rumsey, J., & Michèle, T. (2020). *This Time It 's Different: The Role of Women 's Employment in a Pandemic Recession* (Vol. 92093, Issue 13562).
- Alon, T., Doepke, M., Olmstead-Rumsey, J., & Tertilt, M. (2020). The Impact of Covid-19 on Gender Equality. In *NBER Working Paper Series* (No. 26947; Vol. 53, Issue 9).
- Andrew, A., Cattan, S., Costa Dias, M., Farquharson, C., Kraftman, L., Krutikova, S., Phimister, A., & Sevilla, A. (2020). The Gendered Division of Paid and Domestic Work under Lockdown. *IZA Discussion Paper*, 13500.
- Boca, D. Del, Oggero, N., Profeta, P., & Rossi, M. C. (2020). *Women 's Work , Housework and Childcare , before and during COVID-19 Women 's Work , Housework and Childcare , before and during COVID-19*. 13409.
- Borjas, G. J. (2012). *Labor Economics (Sixth Edition)*.
- BPS. (2019). *Pedoman Pencacah Survei Angkatan Kerja Nasional 2019*. BPS RI.
- BPS. (2020a). *Berita Resmi Statistik: Keadaan Ketenagakerjaan Indonesia Agustus 2020* (Issue 86).
- BPS. (2020b). *Indikator Pasar Tenaga Kerja Indonesia Agustus 2020*. BPS.
- BPS. (2021a). *Berita Resmi Statistik: Pertumbuhan Ekonomi Indonesia Triwulan IV-2020*. In *Www.Bps.Go.Id* (Issue 13). <https://www.bps.go.id/pressrelease/2021/02/05/1811/ekonomi-indonesia-2020-turun-sebesar-2-07-persen--c-to-c-.html>
- BPS. (2021b). *Indikator Pasar Tenaga Kerja Indonesia Februari 2021*.
- Dang, H. A. H., & Nguyen, C. V. (2021). Gender inequality during the COVID-19 pandemic: Income, expenditure, savings, and job loss. *World Development*, 140, 105296. <https://doi.org/10.1016/j.worlddev.2020.105296>
- Dewi, P. M. (2012). Partisipasi Tenaga Kerja Perempuan dalam Meningkatkan Pendapatan Keluarga. *Jurnal Ekonomi Kuantitatif Terapan*, 5(2), 119–124.
- Farré, L., Fawaz, Y., Gonzalez, L., & Graves, J. (2020). *How the COVID-19 Lockdown Affected Gender Inequality in Paid and Unpaid Work in Spain*. 13434.
- Hosmer, D. W., & Lemeshow, S. (1991). *Applied Logistic Regression*. In *Biometrics* (Vol. 47, Issue 4). <https://doi.org/10.2307/2532419>
- Hupkau, C., & Petrongolo, B. (2020). Work, Care and Gender during the COVID-19 Crisis*. *Fiscal Studies*, 41(3), 623–651. <https://doi.org/10.1111/1475-5890.12245>
- ILO. (2016). *Key Indicators Of The Labour*

- Market* (Ninth Edit). ILO.
- ILO. (2017). *World Employment Social Outlook*.
- Kikuchi, S., Kitao, S., & Mikoshiba, M. (2021). Who suffers from the COVID-19 shocks? Labor market heterogeneity and welfare consequences in Japan. *Journal of the Japanese and International Economies*, 59(August 2020), 101117. <https://doi.org/10.1016/j.jjie.2020.101117>
- Kristal, T., & Yaish, M. (2020). Does the coronavirus pandemic level the gender inequality curve? (It doesn't). *Research in Social Stratification and Mobility*, 68(May), 100520. <https://doi.org/10.1016/j.rssm.2020.100520>
- Maryati, S., Bachtiar, N., Ekonomi, F., Andalas, U., Ekonomi, F., Andalas, U., Ekonomi, F., & Andalas, U. (2018). Analisis penawaran tenaga kerja wanita memiliki balita di perdesaan sumatera barat. *Jurnal Ekonomi Dan Pembangunan*, 26(1), 77–88.
- Nicholson, S. (2010). *Intermediate Microeconomics and Its Application* (eleventh e).
- Satgas Covid-19. (2021). *Analisa data COVID-19 Indonesia*. <https://covid19.go.id/p/berita/analisis-data-covid-19-indonesia-update-10-januari-2021>
- Sevilla, A., & Smith, S. (2020). Baby steps: The gender division of childcare during the COVID-19 pandemic. *Oxford Review of Economic Policy*, 36(13302), S169–S186. <https://doi.org/10.1093/oxrep/graa027>

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

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

Bagian naskah berisi:

Judul. Judul tidak melebihi 12 kata dalam Bahasa Indonesia.

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

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

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

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

Metodologi terdiri atas:

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

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

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

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

Publikasi Buku

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