

**PEMODELAN *SEMIPARAMETRIC*
GEOGRAPHICALLY WEIGHTED REGRESSION PADA
KASUS JUMLAH PENDUDUK MISKIN DI PROVINSI
PAPUA MENGGUNAKAN *LINEAR MODEL*
*COREGIONALIZATION***

SKRIPSI



WAHYU DWI RAHMAWATI

H051191004

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

MAKASSAR

2023

**PEMODELAN SEMIPARAMETRIC
GEOGRAPHICALLY WEIGHTED REGRESSION PADA
KASUS JUMLAH PENDUDUK MISKIN DI PROVINSI
PAPUA MENGGUNAKAN LINEAR MODEL
COREGIONALIZATION**

SKRIPSI

**Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains pada
Program Studi Statistika Departemen Statistika Fakultas Matematika dan
Ilmu Pengetahuan Alam Universitas Hasanuddin**

WAHYU DWI RAHMAWATI

H051191004

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

**MAKASSAR
AGUSTUS 2023**

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Pemodelan *Semiparametric Geographically Weighted Regression* pada Kasus Jumlah Penduduk Miskin di Provinsi Papua Menggunakan *Linear Model Coregionalization*

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 18 Agustus 2023



Wahyu Dwi Rahmawati

NIM H051191004

**PEMODELAN SEMIPARAMETRIC
GEOGRAPHICALLY WEIGHTED REGRESSION PADA
KASUS JUMLAH PENDUDUK MISKIN DI PROVINSI
PAPUA MENGGUNAKAN LINEAR MODEL
COREGIONALIZATION**

Disetujui Oleh:

Pembimbing Utama

Pembimbing Pertama



Siswanto, S.Si., M.Si.

NIP. 19920107 201903 1 012



Andi Kresna Jaya, S.Si., M.Si.

NIP. 19731228 20003 1 001

Ketua Program Studi



Dr. Anna Islamiyati, S.Si., M.Si.

NIP. 19770808 200501 2 002

Pada 18 Agustus 2023

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Wahyu Dwi Rahmawati

NIM : H051191004

Program Studi : Statistika

Judul Skripsi : *Pemodelan Semiparametric Geographically Weighted Regression* pada Kasus Jumlah Penduduk Miskin di Provinsi Papua Menggunakan *Linear Model Coregionalization*

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Siswanto, S.Si., M.Si. (.....)
2. Sekretaris : Andi Kresna Jaya, S.Si., M.Si. (.....)
3. Anggota : Sitti Sahriman, S.Si., M.Si. (.....)
4. Anggota : Anisa, S.Si., M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 18 Agustus 2023

KATA PENGANTAR

Assalamu 'alaikum Warahmatullahi Wabaraktuh

Puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* karena atas segala limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa Sallam* beserta keluarga dan para sahabatnya. *Alhamdulillahirobbil'alamin*, berkat nikmat kemudahan dan pertolongan yang diberikan oleh Allah *Subhanahu Wa Ta'ala*, penulis dapat menyelesaikan skripsi yang berjudul “**Pemodelan *Semiparametric Geographically Weighted Regression* pada Kasus Jumlah Penduduk Miskin di Provinsi Papua Menggunakan *Linear Model Coregionalization***” yang disusun sebagai salah satu syarat akademik untuk memperoleh gelar sarjana pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis menyadari bahwa dalam penyelesaian skripsi ini tidak lepas dari bantuan dan dorongan dari berbagai pihak yang senantiasa turut membantu dalam bentuk moril maupun materil sehingga dengan segala keterbatasan kemampuan dan pengetahuan, penulis dapat menyelesaikan skripsi ini. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya kepada Ayahanda **Ali Muhsin** dan Ibunda **Musdalipah** yang telah memberikan dorongan, dukungan, limpahan cinta dan kasih sayang, kesabaran hati, serta dengan sepenuh hati percaya pada penulis terkait penyelesaian tugas akhir ini. Ucapan terima kasih juga penulis haturkan kepada Kakak tersayang **Muhammad Azhar** dan istrinya **Musdalifah Ansar** yang senantiasa mendengarkan segala keluh kesah dan memberi saran kepada penulis. Terima kasih juga kepada adik-adik penulis **Agung Darmawan**, dan **Muhammad Yusuf** yang selalu mengisi keseharian penulis dengan berbagai hal lainnya di sela-sela padatnya kegiatan pribadi penulis.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan dan ketulusan juga penulis ucapkan kepada:

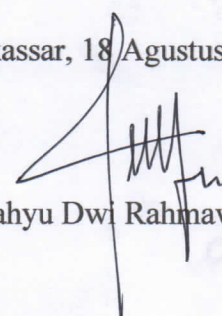
1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.

2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Anna Islamiyati S.Si., M.Si.**, selaku Ketua Departemen Statistika yang dengan penuh kesabaran memberikan arahan, dorongan dan motivasi kepada penulis selama menyelesaikan masa-masa akhir perkuliahan terlebih dalam proses penyelesaian tugas akhir.
4. **Bapak Siswanto, S.Si., M.Si.**, selaku Pembimbing Utama sekaligus Penasehat Akademik penulis yang dengan penuh kesabaran meluangkan waktu dan pemikirannya untuk senantiasa memberikan arahan, dorongan semangat, dan motivasi kepada penulis dari awal perkuliahan, selama penyusunan skripsi hingga akhir masa studi penulis.
5. **Bapak Andi Kresna Jaya, S.Si., M.Si.**, selaku Pembimbing Pertama yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya untuk senantiasa memberikan arahan, dorongan semangat, dan motivasi kepada penulis dari awal hingga selesainya penulisan tugas akhir ini.
6. **Ibu Sitti Sahrinan, S.Si., M.Si.** dan **Ibu Annisa, S.Si., M.Si.**, selaku Tim Penguji yang telah meluangkan waktu untuk memberi motivasi serta kritik dan saran yang membangun kepada penulis dalam penyempurnaan tugas akhir ini.
7. Segenap Dosen Pengajar dan Staf yang telah memberikan ilmu dan kemudahan kepada penulis dalam berbagai hal selama menempuh pendidikan sarjana di Departemen Statistika.
8. Sahabat tercinta penulis, **Andini We Tenri Maharani, Diah Lestari, Fadhilah Febriyanti Najamuddin**, dan **Nurul Hikmah**. Terima kasih telah menjadi sahabat yang mengisi ketegangan selama masa perkuliahan dengan segala jenis hiburan dan canda tawa.
9. Sahabat seperjuangan di Statistika 2019, **Ummul Auliyah Syam, Alya Safira Irtiqa Miolo, Muhammad Yusran, Muh. Nur Iskandar Zulkarnain, Alfiyah Salsa Dila Sabir, A. Ahmad Qeis Tenridapi, Muhammad Rayhan Rifaldi, Vinaya Rifqi Anandari**, dan **Fatimah Utami**. Terima kasih atas kebersamaan, kebahagiaan, kesedihan, serta kebaikannya menjadi sosok guru bagi penulis. Terima kasih telah mengukir kenangan indah bersama penulis selama masa perkuliahan.

10. **Kak Mirnawati, S.Si.** yang telah meluangkan waktunya untuk membantu penulis dalam memahami dan menyelesaikan kendala terkait metode yang digunakan pada tugas akhir ini
11. Teman-teman **Statistika 2019**. Terima kasih atas ilmu, kebersamaan, suka dan duka selama menjalani perkuliahan di Departemen Statistika. Terima kasih sudah menerima kehadiran penulis. Kalian hebat dan luar biasa.
12. **Keluarga Besar Himastat FMIPA Unhas**, terima kasih atas ilmu, pengalaman, dan telah menjadi tempat belajar bagi penulis.
13. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan, partisipasi, dan apresiasi yang diberikan kepada penulis.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini dapat memberikan manfaat untuk berbagai pihak.

Makassar, 18 Agustus 2023


Wahyu Dwi Rahmawati

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIK**

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Wahyu Dwi Rahmawati
NIM : H051191004
Program Studi : Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty- Free Right*)** atas tugas akhir saya yang berjudul:

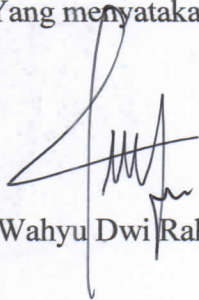
“Pemodelan *Semiparametric Geographically Weighted Regression* pada Kasus Jumlah Penduduk Miskin di Provinsi Papua Menggunakan *Linear Model Coregionalization*”

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar tanggal 18 Agustus 2023.

Yang menyatakan,



(Wahyu Dwi Rahmawati)

ABSTRAK

Semiparametric geographically weighted regression (SGWR) merupakan model regresi yang memuat dua jenis variabel yaitu variabel global dan lokal. Pengelompokan variabel pada penelitian ini memanfaatkan nilai *partial sill* (*psill*) yang diperoleh dari *output linear model coregionalization* (LMC) hasil kombinasi tiga fungsi variogram. Penelitian ini bertujuan untuk mengidentifikasi variabel yang mempengaruhi kemiskinan di Provinsi Papua tahun 2020, diantaranya yaitu angka melek huruf (AMH), angka harapan hidup (AHH), angka partisipasi sekolah (APS), realisasi APBD, jumlah penduduk, pendapatan perkapita, rasio ketergantungan dan tingkat partisipasi angkatan kerja (TPAK). Berdasarkan nilai *psill*, variabel AMH, realisasi APBD, jumlah penduduk, pendapatan perkapita, dan TPAK sebagai variabel global sedangkan variabel AHH, APS dan rasio ketergantungan dikelompokkan sebagai variabel lokal. Kelima variabel global memberikan pengaruh signifikan terhadap jumlah penduduk miskin sedangkan dari ketiga variabel lokal hanya rasio ketergantungan yang memberi pengaruh signifikan di tiap lokasi dan lainnya tidak signifikan di lokasi tertentu. Hasil analisis menunjukkan bahwa model SGWR1 dengan variabel lokalnya memiliki proporsi *psill non-nugget* di atas 80% lebih sesuai dengan data yang dimiliki dibandingkan model SGWR2 yang mengikutsertakan variabel dengan proporsi 70% sampai 80%. Hal tersebut ditunjukkan oleh nilai AICc model SGWR1 sebesar 76.504 sedangkan model SGWR2 sebesar 81.588.

Kata Kunci: Kemiskinan, LMC, Regresi Spasial, SGWR, Variogram.

ABSTRACT

Semiparametric geographically weighted regression (SGWR) is a regression model that contains two types of variables, namely global and local variability. The grouping of variables in this study utilizes the partial sill (psill) value that obtained from the output of the linear coregionalization model (LMC) resulting from the combination of three variogram functions. This study aims to identify variables that affect poverty in Papua Province in 2020, including literacy rate (LR), life expectancy (LE), school participation rate (SPR), RREB realization, population, per capita income, dependency ratio and labor force participation rate (LFPR). Based on the psill value, LR variable, RREB realization, population, per capita income, and LFPR as global variables while LE, SPR and dependency ratio variables are grouped as local variables. The five global variables have a significant influence on the number of poor people, while of the three local variables, only the dependency ratio has a significant influence in each location and others are not significant in certain locations. The results of the analysis showed that the SGWR1 model with its local variables had a proportion of non-nugget psill above 80% more in accordance with the data owned than the SGWR2 model which included variables with a proportion of 70% to 80%. This is indicated by the AICc value of the SGWR1 model of 76.504 while the SGWR2 model is 81.588.

Keywords: LMC, Poverty, SGWR, Spatial Regression, Variogram.

DAFTAR ISI

HALAMAN SAMPUL	i
DAFTAR ISI	ii
HALAMA	iii
LEMBAR PERNYATAAN KEOTENTIKAN	iii
HALAMAN PENGESAHAN	v
KATA PENGANTAR	vi
ABSTRAK	x
ABSTRACT	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA	5
2.1 Regresi Linier Berganda.....	5
2.2 Pengujian Asumsi Model Regresi	5
2.2.1 Uji Normalitas.....	5
2.2.2 Uji Multikolinieritas	6
2.2.3 Heterokedastisitas	6
2.3 <i>Bandwidth</i>	6
2.4 Pembobot Spasial	7
2.5.1 <i>Fixed Kernel</i>	7
2.5.2 <i>Adaptive Kernel</i>	7
2.5 <i>Geographically Weighted Regression</i>	8
2.6 Variogram.....	9
2.7 <i>Linear Model Coregionalization</i>	9
2.8 <i>Semiparametric Geographically Weighted Regression</i>	11
2.9 Pemilihan Model Terbaik	12

2.10 Uji Signifikansi Parameter	12
2.11 Kemiskinan.....	13
BAB III METODE PENELITIAN	14
3.1 Data	14
3.2 Metode Analisis.....	14
BAB IV HASIL DAN PEMBAHASAN	17
4.1 Deskripsi Data	17
4.2 Pengujian Asumsi.....	18
4.2.1 Heterokedastisitas	18
4.3 Fungsi Pembobot Kernel	18
4.4 Variogram.....	20
4.5 <i>Linear Model Coregionalization</i>	22
4.6 <i>Semiparametric Geographically Weighted Regression</i>	25
4.6.1 Estimasi Parameter SGWR.....	25
4.6.2 Pemilihan Model Terbaik	27
4.7 Uji Signifikansi Parameter	29
BAB V KESIMPULAN DAN SARAN	31
5.1 Kesimpulan.....	31
5.2 Saran.....	32
DAFTAR PUSTAKA	33
LAMPIRAN.....	36

DAFTAR GAMBAR

Gambar 4.1. Peta Sebaran Jumlah Penduduk Miskin di Papua.....	17
Gambar 4.2. Plot Sebaran <i>Simple</i> Variogram Y.....	22
Gambar 4.3. Peta Sebaran Penduduk Miskin Hasil Estimasi.....	28

DAFTAR TABEL

Tabel 3.1 Variabel Respon dan Prediktor	14
Tabel 4.1. Nilai <i>Cross Validation Bandwidth</i>	19
Tabel 4.2. Matriks Pembobot dengan Fungsi <i>Fixed Tricube</i>	20
Tabel 4.3. Nilai Kuadrat Selisih Tiap Pasang Lokasi.....	20
Tabel 4.4. Nilai <i>Simple</i> Variogram Variabel Y	21
Tabel 4.5. Jumlah Kuadrat Galat Kombinasi <i>Linear Model Coregionalization</i> ..	23
Tabel 4.6. Proporsi (%) Keragaman Setiap Variabel Berdasarkan Jarak Spasial	24
Tabel 4.7. Proporsi (%) Pengaruh Spasial Variabel	25
Tabel 4.8. Perbandingan Model GWR dan SGWR.....	27
Tabel 4.9. Nilai t_{hitung} Parameter Global dan Lokal	29

DAFTAR LAMPIRAN

Lampiran 1. <i>Output</i> Uji Heterokedastisitas	37
Lampiran 2. Matriks Pembobot.....	38
Lampiran 3. Parameter GWR.....	40
Lampiran 4. Nilai Variogram Variabel	41
Lampiran 5. Plot LMC Model $Nug(0)+Gau(165)+Exp(587)$	43
Lampiran 6. Matriks Koregionalisasi.....	44
Lampiran 7. Parameter SGWR.....	45

BAB I PENDAHULUAN

1.1 Latar Belakang

Kemiskinan merupakan suatu masalah yang masih harus dihadapi masyarakat sampai sekarang. Telah banyak pembangunan yang dilakukan namun belum mampu menekan peningkatan jumlah penduduk miskin, khususnya di negara-negara berkembang (Azizah et al., 2018). Indonesia menjadi salah satu negara yang tidak dapat dikecualikan mengingat status negara berkembang tersebut, dan hasil publikasi Badan Pusat Statistik (BPS) menunjukkan bahwa Provinsi Papua merupakan provinsi dengan persentase jumlah penduduk miskin tertinggi tiap tahunnya dibandingkan provinsi lainnya di Indonesia. Menurut Dores dan Jolianis (2014) angka melek huruf (AMH) dan angka harapan hidup (AHH) dapat mempengaruhi kemiskinan karena semakin tinggi AMH dan AHH maka semakin tinggi pula mutu SDM yang dimiliki.

Pendapatan perkapita juga dapat menjadi indikator kemakmuran suatu wilayah sedangkan jumlah penduduk yang banyak namun tidak diimbangi dengan pendapatan tinggi dapat meningkatkan kemiskinan (Azizah dkk., 2018). Menurut Hidayat dan Woyanti (2021), APBD diharapkan dapat membantu pertumbuhan ekonomi, sedangkan angka ketergantungan dapat menjadi salah satu indikator untuk menunjukkan kondisi ekonomi suatu wilayah. Selain itu, menurut Supratyoningsih dan Yuliarmi (2022), tingkat partisipasi angkatan kerja juga dianggap mempengaruhi kemiskinan di suatu wilayah. Menurut Karini (2018), angka partisipasi sekolah (APS) merupakan tolak ukur pendidikan di suatu wilayah dan meningkatnya APS dapat menurunkan tingkat kemiskinan. Lebih lanjut, jumlah penduduk miskin di Papua dimodelkan dengan analisis regresi klasik.

Analisis regresi linier klasik merupakan salah satu metode statistika yang digunakan untuk mencari hubungan antara variabel prediktor dengan variabel respon sehingga diperoleh suatu model (Apriyani dkk., 2018). Regresi yang bersifat global mengasumsikan setiap wilayah memiliki karakteristik yang sama sehingga sebuah koefisien regresi dihasilkan untuk semua lokasi penelitian. Namun, penggunaan koefisien global kurang sesuai jika terdapat perbedaan karakteristik pada setiap lokasi penelitian (Yuhan & Sitorus, 2017). Regresi spasial merupakan

sebuah pengembangan dari metode linier klasik yang dikembangkan karena adanya pengaruh lokasi pada data yang dianalisis. Salah satu metode statistika yang dapat digunakan untuk mengatasi masalah spasial adalah *geographically weighted regression* (Lutfiani dkk., 2019).

Metode *geographically weighted regression* (GWR) merupakan pemodelan statistik dengan semua variabel prediktor diasumsikan bervariasi spasial dan menghasilkan parameter berbeda untuk setiap lokasi pengamatan (Utami dkk., 2021). GWR dapat meningkatkan ketelitian hasil regresi karena mempertimbangkan pengaruh geografis lokasi serta mengatasi masalah heterogenitas data yang muncul akibat perbedaan kondisi tiap lokasi (Pratiwi & Ayuningsih, 2023). Namun pada beberapa kasus pemodelan GWR ditemukan adanya variabel yang tidak memiliki pengaruh lokasi dan jika diamati lebih lanjut, dapat ditemukan bahwa parameter tersebut bersifat global atau sama untuk setiap lokasi penelitian (Apriyani dkk., 2018).

Terdapat beberapa metode yang dapat digunakan dalam mengelompokkan variabel lokal dan global, diantaranya yaitu dengan memanfaatkan hasil uji signifikansi parsial parameter GWR, metode tersebut digunakan oleh Khoeriyah pada tahun 2021. Selain itu, Ispriyanti (2017) menggunakan simulasi monte-carlo untuk mengelompokkan variabel lokal dan global. Metode lain yang dapat digunakan untuk klasifikasi variabel adalah *linear model correogionalization* (LMC). LMC merupakan model yang berguna untuk menggambarkan hubungan spasial antar variabel (Mar'ah dkk., 2017). Menurut Ribeiro (2016), dengan penggunaan LMC maka lebih mudah untuk menyatakan proporsi tiap skala spasial terhadap variabilitas keseluruhan model serta dapat menunjukkan proporsi pengaruh geografis masing-masing variabel prediktor dengan variabel respon pada skala spasial yang berbeda.

Metode statistik sebagai alternatif yang dapat digunakan jika analisis regresi memuat variabel lokal dan global adalah model *semiparametric geographically weighted regression* (SGWR). Model SGWR dapat menghasilkan nilai parameter yang sebagian bersifat global dan lainnya bersifat lokal atau parameternya berbeda sesuai lokasi pengamatan. Estimasi parameter lokal SGWR dilakukan dengan metode kuadrat terkecil terboboti sedangkan estimasi parameter global

menggunakan metode kuadrat terkecil (Fotheringham dkk., 2002). Dibandingkan regresi linier klasik ataupun GWR, SGWR sebagai model yang merupakan kombinasi dari regresi linier global dan GWR dapat memberikan informasi terkait variabel apa saja yang berpengaruh di setiap titik lokasi pengamatan maupun secara keseluruhan (Kusumaningati dkk., 2019).

Khoeriyah (2021) meneliti beberapa faktor yang mempengaruhi indeks pembangunan kesehatan masyarakat (IPKM) di Provinsi Sumatera Utara Hasil penelitiannya menunjukkan bahwa model SGWR dapat menjelaskan perubahan nilai IPKM lebih baik dibanding model GWR. Penelitian lainnya menggunakan SGWR oleh Ispriyanti (2017) untuk memodelkan variabel yang mempengaruhi pencemaran udara di Surabaya. Hasil penelitiannya menunjukkan bahwa model SGWR dapat menjelaskan tingkat pencemaran udara di Surabaya dibandingkan model GWR. Berdasarkan hal tersebut, penelitian ini menggunakan SGWR untuk mengidentifikasi variabel yang mempengaruhi jumlah penduduk miskin di Provinsi Papua tahun 2020 dengan pengelompokan variabel lokal dan global didasarkan pada *output* model LMC.

1.2 Rumusan Masalah

Rumusan masalah yang dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana mengelompokkan variabel menggunakan *linear model coregionalization*?
2. Bagaimana mengestimasi parameter model *semiparametric geographically weighted regression* untuk data kemiskinan di Provinsi Papua Tahun 2020?
3. Variabel apa saja yang berpengaruh signifikan terhadap jumlah penduduk miskin di Provinsi Papua tahun 2020 berdasarkan model terbaik yang digunakan?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data kemiskinan tiap kabupaten/kota di Provinsi Papua tahun 2020.
2. Pembentukan matriks pembobot menggunakan fungsi kernel terbaik diantara beberapa fungsi kernel yaitu *Gaussian*, *Bi-Square* dan *Tricube* dengan penggunaan *fixed* dan *adaptive bandwidth*.

3. Variabel dengan persentasi proporsi pengaruh spasial di atas 70% dikelompokkan sebagai variabel lokal
4. Pemilihan model regresi terbaik berdasarkan nilai *corrected akaike information criterion* (AICc) terkecil.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mendapatkan hasil pengelompokan variabel menggunakan *linear model coregionalization*.
2. Mendapatkan estimasi parameter *semiparametric geographically weighted regression* untuk data kemiskinan di Provinsi Papua tahun 2020.
3. Mendapatkan variabel yang berpengaruh signifikan terhadap jumlah penduduk miskin di Provinsi Papua tahun 2020 pada model terbaik yang digunakan.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memberikan informasi tentang pemodelan SGWR menggunakan LMC.
2. Memberikan informasi tentang pembentukan LMC menggunakan model dasar variogram.
3. Sebagai informasi bagi pemerintah khususnya bagi TNP2K maupun masyarakat umum terkait hal-hal yang dapat mempengaruhi penduduk miskin.

BAB II
TINJAUAN PUSTAKA

2.1 Regresi Linier Berganda

Menurut Auqino (2019) regresi yang melibatkan satu variabel respon dan lebih dari satu variabel prediktor disebut regresi linier berganda. Secara umum, model regresi linier berganda dituliskan pada Persamaan (2.1).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p+1} X_{(p+1)} + \varepsilon \quad (2.1)$$

dengan:

Y : matriks berisi nilai variabel respon dengan ukuran $(n \times 1)$

X : matriks berisi nilai variabel prediktor dengan ukuran $n \times (p + 1)$

β : matriks parameter berukuran $(p + 1) \times 1$

ε : vektor galat dengan ukuran $(n \times 1)$

Persamaan (2.1) dapat dituliskan dalam bentuk matriks berikut:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1(p+1)} \\ 1 & X_{21} & \dots & X_{2(p+1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{n(p+1)} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p+1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Salah satu metode yang dapat digunakan untuk mengestimasi parameter regresi linier berganda adalah metode kuadrat terkecil. Metode ini berfungsi untuk meminimumkan jumlah kuadrat galat.

2.2 Pengujian Asumsi Model Regresi

2.2.1 Uji Normalitas

Uji *Kolmogorov-Smirnov* digunakan untuk mengetahui bahwa sampel berasal dari data populasi dengan distribusi tertentu. Pengujian ini dimaksudkan untuk membandingkan distribusi data yang dimiliki dengan distribusi normal. Statistik uji *Kolmogorov-Smirnov* dapat dihitung menggunakan Persamaan (2.2).

$$L = |F(z_i) - S(z_i)| \quad (2.2)$$

dengan:

$F(z_i)$: Probabilitas teoritis

$S(z_i)$: Probabilitas kumulatif empiris

$$z_i = \frac{x_i - \bar{x}}{sd}$$

2.2.2 Uji Multikolinieritas

Multikolinieritas merupakan kondisi adanya korelasi antara variabel prediktor. Nilai *variance inflation factor* (VIF) digunakan untuk mendeteksi adanya multikolinieritas pada regresi linier yang melibatkan lebih dari dua variabel prediktor. Nilai VIF lebih besar dari 10 mengindikasikan terjadinya multikolinieritas. Nilai VIF untuk koefisien regresi ke-*k* dapat dihitung menggunakan Persamaan (2.3).

$$VIF_k = \frac{1}{1 - R_k} \tag{2.3}$$

dengan $k = 1, 2, 3, \dots, p$ dan nilai R_k adalah koefisien determinasi antara x_k dengan variabel prediktor lainnya (Sriningsih dkk., 2018).

2.2.3 Heterokedastisitas

Asumsi heterokedastisitas berarti terdapat ragam galat yang tidak sama. *Breusch-Pagan* merupakan sebuah uji statistik untuk mendeteksi heterokedastisitas spasial. Berikut Persamaan (2.4) digunakan untuk menghitung statistik uji (Khariyani dkk., 2022):

$$BP = \frac{1}{2} \mathbf{f}^t \mathbf{Z} (\mathbf{Z}^t \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{f} \tag{2.4}$$

dengan:

$$f_i = \frac{e_i^2}{\sigma^2} - 1$$

BP : nilai uji *Breusch-Pagan*

e_i : galat untuk pengamatan ke-*i* ; $i = 1, 2, \dots, n$

Z : matriks *X* berukuran $n \times (p + 1)$ dari setiap pengamatan terstandarisasi

σ^2 : ragam dari e_i

2.3 Bandwidth

Bandwidth merupakan parameter pemulus estimasi menggunakan pendekatan kernel. Nilai *bandwidth* yang terlalu kecil akan menyebabkan fungsi yang diestimasi menjadi kasar sehingga hubungan variansinya tinggi, sebaliknya jika nilai *bandwidth* terlalu besar menyebabkan fungsi yang diestimasi sangat mulus sehingga hubungan variansinya rendah. Oleh karena itu, diperlukan pemilihan nilai *bandwidth* yang optimal (Saputra & Listyani, 2016). *Cross validation* (CV) merupakan salah satu kriteria untuk menentukan nilai *bandwidth*

yang optimum. Nilai *bandwidth* yang optimum akan menghasilkan CV yang kecil, nilai CV dapat dihitung menggunakan Persamaan (2.5).

$$CV = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(b)]^2 \quad (2.5)$$

dengan nilai $\hat{y}_{\neq i}(b)$ merupakan nilai dugaan y_i namun pengamatan pada lokasi ke- i dikecualikan dengan nilai b merupakan nilai *bandwidth* yang diuji.

2.4 Pembobot Spasial

Pembobot bergantung pada jarak antar titik lokasi pengamatan, semakin jauh jarak lokasi semakin kecil nilai bobotnya bahkan bisa mencapai nilai nol. Pembobot merupakan matriks dengan elemen diagonalnya adalah nilai fungsi pembobot untuk setiap titik lokasi pengamatan. Salah satu metode pembobot yang biasa digunakan adalah fungsi kernel (Lutfiani dkk., 2019). Terdapat dua jenis fungsi kernel, yaitu *fixed kernel* atau fungsi kernel tetap dan *adaptive kernel* atau fungsi kernel adaptif (Tizona dkk., 2017).

2.5.1 Fixed Kernel

Fixed kernel memiliki *bandwidth* (b) yang sama di setiap titik lokasi pengamatan dengan d_{ij} merupakan jarak antar lokasi pengamatan. Persamaan (2.6) digunakan untuk menghitung nilai matriks pembobot dengan *bandwidth* tetap (Tizona dkk., 2017).

a. Fungsi *Kernel Gaussian*

$$w_{ij} = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{b} \right)^2 \right] \quad (2.6.a)$$

b. Fungsi *Kernel Bi-Square*

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b} \right)^2 \right]^2 & , \text{jika } d_{ij} \leq b \\ 0 & , \text{lainnya} \end{cases} \quad (2.6.b)$$

c. Fungsi *Kernel Tricube*

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b} \right)^3 \right]^3 & , \text{jika } d_{ij} \leq b \\ 0 & , \text{lainnya} \end{cases} \quad (2.6.c)$$

2.5.2 Adaptive Kernel

Fungsi *adaptive kernel* serupa dengan *fixed kernel* namun memiliki *bandwidth* (b) yang berbeda di setiap titik lokasi pengamatan. Persamaan (2.7) digunakan untuk menghitung pembobot *bandwidth* adaptif (Tizona dkk., 2017).

a. Fungsi *Kernel Gaussian*

$$w_{ij} = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{b_i} \right)^2 \right] \quad (2.7.a)$$

b. Fungsi *Kernel Bi-Square*

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b_i} \right)^2 \right]^2 & , \text{jika } d_{ij} \leq b_i \\ 0 & , \text{lainnya} \end{cases} \quad (2.7.b)$$

c. Fungsi *Kernel Tricube*

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b_i} \right)^3 \right]^3 & , \text{jika } d_{ij} \leq b_i \\ 0 & , \text{lainnya} \end{cases} \quad (2.7.c)$$

dengan $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$ merupakan jarak *euclidean* antara titik lokasi pengamatan ke- i dengan titik lokasi pengamatan ke- j , dan nilai b_i merupakan nilai *bandwidth* di lokasi ke- i .

2.5 Geographically Weighted Regression

Geographically weighted regression (GWR) merupakan perluasan spasial dari regresi linier berganda, sehingga regresi tidak terbatas pada variabel global, namun juga mempertimbangkan variabel lokal. Berikut Persamaan (2.8) yang digunakan untuk pemodelan GWR (Manganelli dkk., 2014):

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \quad (2.8)$$

dengan:

y_i : nilai variabel respon di tiap wilayah ; $i = 1, 2, 3, \dots, n$

$\beta_0(u_i, v_i)$: koefisien intersep lokasi i

$\beta_k(u_i, v_i)$: nilai parameter prediktor ke- k lokasi i ; $k = 1, 2, 3, \dots, p$

x_{ik} : nilai variabel prediktor ke- k lokasi i

ε_i : galat pengamatan ke- i

Estimasi parameter pada GWR pada Persamaan (2.9) menggunakan metode kuadrat terkecil terboboti (Lu dkk., 2017):

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^t \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}(u_i, v_i) \mathbf{y} \quad (2.9)$$

Dengan nilai \mathbf{X} merupakan nilai variabel prediktor berukuran $n \times (p + 1)$ dan matriks \mathbf{W} merupakan matriks diagonal berukuran $n \times n$ yang menunjukkan bobot geografis untuk tiap lokasi pengamatan.

2.6 Variogram

Pendugaan pada data spasial menggunakan sesuatu yang dapat menggambarkan, memodelkan dan menghitung varians spasial antara variabel x_i dan x_j yang disebut variogram (Rozalia dkk., 2016). Isaaks dan Srivasta (1989) menyatakan bahwa variogram dapat menggambarkan hubungan spasial antar peubah. Variogram didefinisikan oleh Persamaan (2.10) dan (2.11).

$$\gamma_Y(h) = \frac{1}{2N(h)} \sum_{(i,j)=1}^{N(h)} (y_i - y_j)^2 \quad (2.10)$$

$$\gamma_X(h) = \frac{1}{2N(h)} \sum_{(i,j)=1}^{N(h)} (x_i - x_j)^2 \quad (2.11)$$

dengan:

y_i : nilai variabel respon pada lokasi ke- i , $i = 1, 2, 3, \dots, n$

y_j : nilai variabel respon pada lokasi ke- j , $j = 1, 2, 3, \dots, n$

x_i : nilai variabel prediktor pada lokasi ke- i

x_j : nilai variabel prediktor pada lokasi ke- j ; $i \neq j$

$N(h)$: banyak pasang titik dalam radius h

Terdapat tiga parameter variogram diantaranya adalah *nugget effect* atau nilai variogram pada jarak sekitar nol, *sill* atau nilai varian data ketika nilai variogram mencapai nilai stabil, dan *range* (a) adalah jarak ketika variogram mencapai *sill* (Rozalia dkk., 2016). Terdapat beberapa pemodelan variogram terstandarisasi yang sering digunakan, diantaranya yaitu (Isaaks & Srivasta, 1989):

$$\text{Nugget} \quad :g(h) = \begin{cases} 0 & \text{jika } h = 0 \\ 1 & \text{lainnya} \end{cases} \quad (2.12)$$

$$\text{Spherical} \quad :g(h) = \begin{cases} 1,5 \left(\frac{h}{a}\right) - 0,5 \left(\frac{h}{a}\right)^3 & \text{jika } h \leq a \\ 1 & \text{lainnya} \end{cases} \quad (2.13)$$

$$\text{Exponential} \quad :g(h) = 1 - \exp\left(-\frac{h}{a}\right) \quad (2.14)$$

$$\text{Gaussian} \quad :g(h) = 1 - \exp\left(-\frac{h^2}{a^2}\right) \quad (2.15)$$

dengan nilai h dan a masing-masing melambangkan nilai jarak antar lokasi sampel dan *range* atau jarak saat nilai variogram mencapai *sill*.

2.7 Linear Model Coregionalization

Linear model coregionalization (LMC) merupakan alat yang berguna untuk menggambarkan hubungan spasial antara prediktor dan respon. LMC terdiri dari

variogram dan variogram silang dari dua atau lebih variabel. Variogram akan mendeskripsikan karakteristik tiap variabel sedangkan variogram silang mendeskripsikan tiap pasang variabel. Berikut Persamaan (2.16) yang digunakan untuk menyusun LMC (Mar'ah dkk., 2017):

$$\gamma_{kk'}(h) = \sum_{s=0}^t c_{kk',s} g_s(h_s) \quad (2.16)$$

dengan:

$\gamma_{kk'}$: varians variabel prediktor ke k dan k'

$c_{kk',s}$: koefisien variogram untuk tiap model variogram

g_s : model variogram ke-s

h_s : jarak untuk model variogram model variogram ke-s

dengan $k, k' = 1, 2, \dots, p$ dan nilai p merupakan banyaknya variabel prediktor.

LMC pada penelitian ini terbentuk dari pengaruh *nugget* dan dua model semivariogram dasar lainnya sebagai berikut :

$$\gamma_{kk'}(h) = c_{kk',0} g_0(h_0) + c_{kk',1} g_1(h_1) + c_{kk',2} g_2(h_2)$$

dengan:

$c_{kk',0}$: koefisien variogram model *nugget*

$c_{kk',1}$: koefisien variogram untuk model variogram dasar pertama

$c_{kk',2}$: koefisien variogram untuk model semivariogram dasar kedua

g_0 : variogram model *nugget*

g_1 : model variogram dasar pertama

g_2 : model variogram dasar kedua

h_0 : jarak untuk model variogram *nugget*

h_1 : jarak untuk model variogram dasar pertama

h_2 : jarak untuk model variogram dasar kedua

Model variogram dasar yang digunakan sama untuk setiap model variogram maupun variogram silang, selanjutnya matriks LMC disusun berdasarkan Persamaan (2.17).

$$\Gamma = \begin{bmatrix} \gamma_{11}(h) & \dots & \gamma_{1p'}(h) \\ \vdots & \ddots & \vdots \\ \gamma_{k1}(h) & \dots & \gamma_{pp}(h) \end{bmatrix} = \sum_{s=0}^p \mathbf{C}_s g_s(h_s) \quad (2.17)$$

dengan matriks \mathbf{C}_s merupakan matriks definit positif berukuran $p \times p$ dan $g_s(h_s)$ merupakan fungsi variogram terstandarisasi yang digunakan. Matriks \mathbf{C}_s berisi

nilai varian-kovarian parsial untuk jarak spasial berbeda dengan elemen diagonalnya merupakan koefisien model variogram dan elemen lainnya merupakan koefisien model variogram silang. Nilai variogram yang berada pada diagonal matriks koregionalisasi untuk skala non-nugget kemudian digunakan sebagai landasan pengelompokan variabel. Variabel dengan proporsi spasial tinggi akan dikelompokkan sebagai variabel lokal dan sebaliknya, variabel dengan proporsi spasial rendah dikelompokkan sebagai variabel lokal (Ribeiro et al., 2016).

2.8 *Semiparametric Geographically Weighted Regression*

Pada situasi tertentu, tidak semua koefisien regresi terboboti bersifat spasial (koefisien berbeda untuk lokasi berbeda). Tidak dapat diabaikan kemungkinan adanya variabel yang berpengaruh secara global (sama untuk setiap lokasi), sehingga untuk mengatasi dua koefisien dengan sifat berbeda dapat digunakan model *semiparametric geographically weighted regression* (SGWR). Berikut Persamaan (2.20) yang digunakan untuk memodelkan SGWR (Fotheringham dkk., 2002):

$$y_i = \sum_{g=1, k_g} \beta_g x_{ig}(g) + \sum_{l=1, k_l} \beta_l(u_i, v_i) x_{il}(l) + \varepsilon_i \quad (2.18)$$

dengan (u_i, v_i) adalah koordinat lokasi yang diobservasi, β_g adalah koefisien untuk variabel global ke- g dan $\beta_l(u_i, v_i)$ adalah koefisien untuk variabel lokal ke- l pada lokasi ke- i . Estimasi parameter lokal dan global SGWR masing-masing dituliskan pada Persamaan (2.19) dan (2.20).

$$\hat{\beta}_l(u_i, v_i) = (\mathbf{X}_l^t \mathbf{W}(u_i, v_i) \mathbf{X}_l)^{-1} \mathbf{X}_l^t \mathbf{W}(u_i, v_i) (\mathbf{y} - \mathbf{X}_g \beta_g) \quad (2.19)$$

$$\hat{\beta}_g = (\mathbf{X}_g^t (\mathbf{I} - \mathbf{S}_l)^t (\mathbf{I} - \mathbf{S}_l) \mathbf{X}_g)^{-1} \mathbf{X}_g^t (\mathbf{I} - \mathbf{S}_l)^t (\mathbf{I} - \mathbf{S}_l) \mathbf{y} \quad (2.20)$$

dengan:

$\beta_l(u_i, v_i)$: matriks parameter lokal pada lokasi ke- i berukuran $(k_l + 1) \times 1$

\mathbf{X}_l : matriks variabel lokal berukuran $n \times (k_l + 1)$

$\mathbf{W}(u_i, v_i)$: matriks dengan pembobot ke- i berada pada diagonal ; $i = 1, 2, 3, \dots, n$

\mathbf{y} : matriks variabel respon berukuran $n \times 1$

β_g : matriks parameter global berukuran $k_g \times 1$

\mathbf{X}_g : matriks variabel global berukuran $n \times k_g$

$\mathbf{S}_l = \mathbf{X}_l (\mathbf{X}_l^t \mathbf{W}(u_i, v_i) \mathbf{X}_l)^{-1} \mathbf{X}_l^t \mathbf{W}(u_i, v_i)$

2.9 Pemilihan Model Terbaik

Salah satu metode yang dapat digunakan untuk pemilihan model terbaik yaitu *corrected akaike information criterion* (AICc). Nilai AICc merepresentasikan ukuran jarak informasi antara distribusi model yang diuji dengan distribusi sebenarnya, sehingga semakin kecil nilainya maka semakin baik model tersebut. Nilai AICc dapat dihitung menggunakan Persamaan (2.22) berikut (Fotheringham dkk., 2002):

$$AICc = 2n \log_e(\hat{\sigma}) + n \log_e(2\pi) + n \left\{ \frac{n + tr(\mathbf{S})}{n - 2 - tr(\mathbf{S})} \right\} \quad (2.21)$$

dengan:

$$\hat{\sigma}^2 = \frac{SSE}{n}$$

$$\mathbf{S} = \mathbf{S}_l + (\mathbf{I} - \mathbf{S}_l)\mathbf{X}_g(\mathbf{X}_g^t(\mathbf{I} - \mathbf{S}_l)^t(\mathbf{I} - \mathbf{S}_l)\mathbf{X}_g)^{-1}\mathbf{X}_g^t(\mathbf{I} - \mathbf{S}_l)^t(\mathbf{I} - \mathbf{S}_l)$$

2.10 Uji Signifikansi Parameter

Menurut Pratiwi dan Ayuningsih (2023) estimasi parameter menggunakan metode SGWR menghasilkan dua jenis parameter dengan variasi berbeda sehingga pengujian signifikansi parameter dilakukan secara terpisah. Uji signifikansi parsial parameter global SGWR dilakukan menggunakan Persamaan (2.22)

$$t_g = \frac{\hat{\beta}_k}{\hat{\sigma} \sqrt{g_{kk}}} \quad (2.22)$$

nilai g_{kk} merupakan elemen diagonal ke- k dari matriks $\mathbf{G}\mathbf{G}^t$, dengan:

$$\mathbf{G} = (\mathbf{X}_g^t(\mathbf{I} - \mathbf{S}_l)^t(\mathbf{I} - \mathbf{S}_l)\mathbf{X}_g)^{-1}\mathbf{X}_g^t(\mathbf{I} - \mathbf{S}_l)^t(\mathbf{I} - \mathbf{S}_l)$$

$$\hat{\sigma}^2 = \frac{\mathbf{y}^t(\mathbf{I} - \mathbf{S})^t(\mathbf{I} - \mathbf{S})\mathbf{y}}{tr(\mathbf{I} - \mathbf{S})^t(\mathbf{I} - \mathbf{S})}$$

Berbeda dengan parameter global, pengujian signifikansi parameter lokal model SGWR menggunakan Persamaan (2.23).

$$t_l = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma} \sqrt{m_{kk}}} \quad (2.23)$$

nilai m_{kk} merupakan elemen diagonal ke- k dari matriks $\mathbf{M}\mathbf{M}^t$, dengan:

$$\mathbf{M} = (\mathbf{X}_l^t\mathbf{W}(u_i, v_i)\mathbf{X}_l)^{-1}\mathbf{X}_l^t\mathbf{W}(u_i, v_i)(\mathbf{I} - \mathbf{X}_g\mathbf{G})$$

$$\hat{\sigma}^2 = \frac{\mathbf{y}^t(\mathbf{I} - \mathbf{S})^t(\mathbf{I} - \mathbf{S})\mathbf{y}}{tr(\mathbf{I} - \mathbf{S})^t(\mathbf{I} - \mathbf{S})}$$

2.11 Kemiskinan

Kemiskinan merupakan keadaan seseorang atau sekelompok orang tidak mampu memenuhi pendidikan, pangan, kesehatan, sanitasi, dan hal lainnya untuk mempertahankan kehidupan (Hardinandar, 2019). Sebagian besar negara di dunia menjadikan kemiskinan sebagai masalah utama negara, terlebih bagi negara yang masih berstatus negara berkembang, salah satunya di Indonesia (Priseptian & Primandhana, 2022). Berdasarkan data publikasi Badan Pusat Statistik (BPS), persentase penduduk miskin di Provinsi Papua selalu melebihi 25% sekaligus menjadi provinsi dengan persentase tertinggi. Menurut BPS, berikut adalah pengertian dari variabel-variabel prediktor yang akan digunakan :

1. Angka melek aksara/huruf (AMH) penduduk usia ≥ 15 adalah perbandingan jumlah penduduk berusia ≥ 15 tahun yang dapat membaca dan menulis kalimat sederhana dengan huruf latin dan atau huruf lainnya terhadap jumlah penduduk berusia ≥ 15 tahun.
2. Angka harapan hidup (AHH) pada waktu lahir merupakan rata-rata perkiraan banyak tahun yang dapat ditempuh seseorang selama hidup.
3. Angka partisipasi sekolah usia 16-18 tahun adalah perbandingan penduduk berusia 16-18 tahun yang sedang mengenyam pendidikan formal terhadap jumlah penduduk berusia 16-18 tahun.
4. Realisasi anggaran pendapatan dan belanja daerah adalah perhitungan APBD kabupaten/kota tiap tahun anggaran.
5. Jumlah penduduk tiap kab/kota merupakan banyaknya orang yang berdomisili di suatu kab/kota minimal satu tahun atau kurang dari satu tahun dengan tujuan menetap.
6. Pendapatan perkapita adalah jumlah pendapatan yang diterima seluruh rumah tangga keluarga dibagi dengan jumlah penduduk pertengahan tahun.
7. Rasio ketergantungan adalah perbandingan jumlah penduduk usia 0-14 tahun ditambah penduduk berusia 65 tahun ke atas dengan jumlah penduduk usia 15-64 tahun.
8. Tingkat partisipasi angkatan kerja adalah persentase banyaknya angkatan kerja terhadap jumlah penduduk usia 10 tahun ke atas.