GEOGRAPHICALLY WEIGHTED REGRESSION PRINCIPAL COMPONENT ANALYSIS (GWRPCA) PADA IPKM DI KALIMANTAN BARAT

IKHWAN MA'RUF
H1091191043

SKRIPSI

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2024

GEOGRAPHICALLY WEIGHTED REGRESSION PRINCIPAL COMPONENT ANALYSIS (GWRPCA) PADA IPKM DI KALIMANTAN BARAT

IKHWAN MA'RUF
H1091191043

Skripsi
Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Statistika pada Program Studi Statistika

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2024

LEMBAR PENGESAHAN

Judul Tugas Akhir	: Geographically Weighted Regression Principal Component Analysis (GWRPCA) Pada IPKM Di Kalimantan Barat
Nama Mahasiswa	: Ikhwan Ma'ruf
NIM	: H1091191043
Jurusan/Program Studi	: Matematika/Statistika
Tanggal Lulus	: 19 November 2024
SK Pembimbing	: No. 4823/UN22.8/TD.06/22 Desember 2023
SK Penguji	: No. 3723/UN22.8/TD.06/14 November 2024

Dosen Pembimbing

Pembimbing I	Pembimbing II
Hendra Perdana, M.Sc	Shantika Martha, M.Si
NIP 198810102019031020	NIP 198403082008122003

Dosen Penguji

Ketua Penguji	Anggota Penguji
Nurfitri Imro'ah, M.Si	Siti Aprizkiyandari, M.Si
NIP 198907182019032021	NIP 199404282019032035
Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Hendra Perdana, M.Sc	Shantika Martha, M.Si
NIP 198810102019031020	NIP 198403082008122003

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

Prof. Dr. Gusrizal, S.Si., M.Si
NIP 197108022000031001

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama : Ikhwan Ma'ruf

NIM : H1091191043

Program Studi/ Jurusan : Matematika/Statistika

Fakultas : Matematika Dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyataan ini untuk digunakan sebagaimana mestinya.

Pontianak, 19 November 2024

Ikhwan Ma'ruf
NIM H1091191043

***Geographically Weighted Regression Principal Component Analysis* (GWRPCA) pada IPKM di Kalimantan Barat**

Abstrak

Geographically Weighted Regression Principal Component Analysis (GWRPCA) merupakan gabungan dari metode *Geographically Weighted Regression* (GWR) dan *Principal Component Analysis* (PCA). Sehingga metode GWRPCA sangat cocok digunakan untuk menganalisis Indeks Pembangunan Kesehatan Masyarakat (IPKM) di Kalimantan Barat. Kesehatan balita, kesehatan reproduksi, pelayanan kesehatan, perilaku kesehatan, penyakit tidak menular, penyakit menular dan kesehatan lingkungan adalah variabel independen yang digunakan. Penelitian ini dimulai dengan proses analisis pemodelan regresi linear berganda. Selanjutnya, analisis GWRPCA digunakan untuk menemukan nilai koefisien determinasi global dan lokal. Hasil penelitian menunjukkan bahwa nilai koefisien determinasi (R^2) global sebesar 96,01%. Di sisi lain, untuk model GWRPCA 14 model ditemukan nilai koefisien determinasi lokal pada setiap lokasi. Sambas memiliki koefisien determinasi lokal tertinggi sebesar 96,06%, dan Kapuas Hulu memiliki koefisien determinasi lokal terkecil sebesar 95,81%.

Kata kunci: IPKM, Multikolinieritas, Spasial.

Geographically Weighted Regression Principal Component Analysis (GWRPCA) on IPKM in West Kalimantan

Abstract

Geographically Weighted Regression Principal Component Analysis (GWRPCA) is a combination of Geographically Weighted Regression (GWR) and Principal Component Analysis (PCA). Therefore, the GWRPCA method is very suitable for analyzing the Public Health Development Index (IPKM) in West Kalimantan. Variables used include child health, reproductive health, health services, health behavior, non-communicable diseases, communicable diseases, and environmental health. This research begins with the process of multiple linear regression modeling analysis. Subsequently, GWRPCA analysis is used to find the global and local coefficient of determination values. The results indicate that the global coefficient of determination (R^2) is 96.01%. On the other hand, in the GWRPCA model, 14 models were found with local coefficient of determination values at each location. Sambas has the highest local coefficient of determination at 96.06%, while Kapuas Hulu has the smallest local coefficient of determination at 95.81%.

Keywords: *IPKM, Multicollinearity, Spatial.*

PRAKATA

Bismillahirrahmanirrahim

Assalamu'alaikum Wr. Wb.

Puji syukur kehadirat ALLAH SWT atas segala nikmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "***Geographically Weighted Regression Principal Component Analysis* (GWRPCA) Pada IPKM Di Kalimantan Barat**". Tujuan penulisan skripsi ini adalah untuk memenuhi syarat memperoleh gelar Sarjana Statistika (S.Stat) bagi mahasiswa Program Studi S1 Statistika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura.

Tersusunnya skripsi ini tidak lepas dari bantuan, arahan, dan petunjuk dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kedua orang tua tercinta Akhmad Rudi dan Rapiqah, kakak-kakakku, Nurul Hijayanti, Ety Wahyuni dan Azarini Rizqia, serta keluarga penulis yang telah banyak memberikan do'a, dan dorongan serta motivasi dalam penulisan skripsi ini.
2. Bapak Yudhi, S.Si.,M.Si., selaku Dosen Pembimbing Akademik yang telah memberikan bimbingan, arahan, dorongan, serta motivasi untuk penulis selama perkuliahan.
3. Bapak Hendra Perdana, M.Sc., selaku Dosen Pembimbing Pertama yang telah membimbing dan memberikan arahan selama penulisan skripsi ini.
4. Ibu Shantika Martha, M.Si selaku Dosen Pembimbing Kedua yang telah membimbing dan memberikan arahan selama penulisan skripsi ini.
5. Ibu Nurfitri Imro'ah, M.Si selaku dosen penguji pertama yang telah memberikan masukkan dan saran kepada penulis.
6. Ibu Siti Aprizkiyandari, M.Si., selaku dosen penguji kedua yang telah memberikan masukkan dan saran kepada penulis.
7. Rekan-rekan mahasiswa Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura Pontianak Angkatan 2019 yang telah menemani, memberi semangat, kritik, dan motivasi selama proses perkuliahan, penelitian dan penulisan skripsi ini,

Semoga amal baiknya diterima ALLAH SWT dan mendapat pahala yang setimpal dari-NYA. Aamiin.

Akhir kata, semoga skripsi ini bermanfaat bagi penulis dan semua pihak yang membutuhkannya.

Wabillahi taufik wal hidayah

Wassalamu'alaikum Wr. Wb

Pontianak, 19 November 2024

Ikhwan Ma'ruf

DAFTAR ISI

Abstrak ..... **iv**
Abstract ..... **v**
PRAKATA ..... **vi**
DAFTAR ISI ..... **viii**
DAFTAR GAMBAR ..... **x**
DAFTAR TABEL ..... **xi**
DAFTAR LAMPIRAN ..... **xii**
DAFTAR SIMBOL ..... **xiii**
BAB I PENDAHULUAN **14**
1.1 Latar Belakang 14
1.2 Rumusan Masalah 16
1.3 Tujuan Penelitian 16
1.4 Batasan Masalah 17
1.5 Tinjauan Pustaka 17
1.6 Metodologi Penelitian 19
BAB II LANDASAN TEORI **20**
2.1 Indeks Pembangunan Kesehatan Masyarakat (IPKM) 20
2.2 Definisi Operasional Indikator IPKM 20
2.3 Analisis Faktor 21
2.4 Regresi Linear 21
2.5 Pengujian Asumsi Klasik 22
BAB III *GEOGRAPHICALLY WEIGHTED REGRESSION PRINCIPAL COMPONENT ANALYSIS* (GWRPCA) **25**
3.1 Analisis Data Spasial 25
3.2 *Geographically Weighted Regression* (GWR) 26

3.3 Estimasi Parameter Model GWR 26
3.4 Pemilihan Bandwidth Optimum 28
3.5 Pengujian Model GWR 28
3.6 *Geographically Weighted Regression Principal Component Analysis* (GWRPCA) 30
3.7 Pembobotan Model GWRPCA 31
3.8 Estimasi Parameter Model GWRPCA 32
3.9 Pengujian Hipotesis Model GWRPCA 33
3.10 Koefisien Determinasi (R**2**) 34
BAB IV PEMBAHASAN 35
4.1 Deskripsi Data 35
4.2 Statistik Deskriptif 36
4.3 Analisis Faktor 38
4.4 Pemodelan Regresi Linear Berganda 39
4.5 Uji Asumsi Regresi Linear 40
4.5.1 Uji Normalitas 40
4.5.2 Uji Autokorelasi 40
4.5.3 Uji Multikolinieritas 40
4.5.4 Uji Heteroskedastisitas 41
4.6 Pengujian Pengaruh Heterogenitas Spasial 41
4.7 Menghitung Jarak Euclidean 41
4.8 Menghitung Matriks Pembobot 42
4.9 Model GWRPCA 42
4.10 Estimasi Parameter Model GWRPCA 43
4.11 Koefisien Determinasi ($\boldsymbol{R2}$) 43
BAB V KESIMPULAN 45
5.1 Kesimpulan 45
5.2 Saran 45
DAFTAR PUSTAKA 47
LAMPIRAN 49

DAFTAR GAMBAR

Gambar 1 Flowchart pemodelan indeks pembangunan kesehatan masyarakat dengan metode *Geographically Weighted Regression Principal Component Analysis* (GWRPCA) 19

DAFTAR TABEL

Tabel 4. 1 Variabel Laten dan Indikator Penyusun Variabel Laten 35
Tabel 4. 2 Statistik Deskriptif Data Penelitian.....36
Tabel 4. 3 Model Regresi Linier Berganda.....39
Tabel 4. 4 Nilai VIF Model Regresi Linear.....40
Tabel 4. 5 Nilai Cross Validation (CV) dan Bandwidth fungsi pembobot 42
Tabel 4. 6 Hasil Parameter Model GWRPCA.....43

DAFTAR LAMPIRAN

Lampiran 1 Data Indeks Pembangunan Kesehatan Masyarakat dan Faktor-Faktor Yang Mempengaruhinya 49

Lampiran 2 Lanjutan Data Indeks Pembangunan Kesehatan Masyarakat dan Faktor-Faktor Yang Mempengaruhinya 50

Lampiran 3 Data setelah di Analisis Faktor 51

Lampiran 4 Menghitung Euclidean 52

Lampiran 5 Matriks Pembobot 53

Lampiran 6 Syntax Perhitungan Pada Software R Studio 54

DAFTAR SIMBOL

n	: Jumlah sampel
y_i	: Variabel dependen pengamatan ke-*i*
x_{ik}	: Variabel independen ke-*k* pengamatan ke-*i*
β_k	: Koefisien regresi dari variabel x_k
β_0	: *Intercept*
ε_i	: *Error* random ke-*i*
$\boldsymbol{y}$	: Vektor variabel dependen berukuran $n \times 1$
$\boldsymbol{X}$	: Matriks variabel independen berukuran $n \times (p + 1)$
$\boldsymbol{\beta}$	: Vektor parameter berukuran $(p + 1) \times 1$
$\boldsymbol{\varepsilon}$	: Vektor *error* berukuran $n \times 1$
BP	: Nilai uji *Breusch-Pagan*
e_i	: Residual untuk pengamatan ke-*i*
$\boldsymbol{Z}$	: Matriks berukuran $n \times (p + 1)$ yang sudah distandarisasi untuk setiap pengamatan
σ^2	: Ragam dari e_i
$\beta_0(u_i, v_i)$	: Konstanta/*intercept* pada pengamatan ke-*i*
(u_i, v_i)	: menyatakan titik koordinat (*latitude, longitude*) lokasi ke- i
$\beta_k(u_i v_i)$	: Koefisien regresi variabel independen ke-*k* di lokasi pengamatan ke-*i*
$\boldsymbol{W}(u_i, v_i)$	: Matriks pembobot diagonal berukuran ($n \times n$)
h	: Bandwidth
$\boldsymbol{I}$	: Matriks identitas berukuran n × n
$\boldsymbol{L}$	: Matriks proyeksi dari model GWR

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kesehatan sangat penting bagi produktivitas individu. Setiap orang memerlukan pola hidup sehat agar dapat terus hidup. Menurut Undang-Undang Republik Indonesia Nomor 36 tahun 2009, kesehatan didefinisikan sebagai kondisi sehat secara fisik, mental, spiritual, dan sosial, yang memungkinkan individu untuk hidup produktif baik secara sosial maupun ekonomi. Kesehatan adalah salah satu elemen kesejahteraan yang harus diwujudkan dan merupakan hak asasi setiap individu. Hal ini sejalan dengan cita-cita bangsa Indonesia yang tercantum dalam Pancasila dan Undang-Undang Dasar Tahun 1945. Dengan demikian, dapat disimpulkan bahwa kesehatan itu bersifat holistik, mencakup tidak hanya aspek fisik tetapi juga jiwa, sosial, dan ekonomi (Sulistiarini, 2018).

Kondisi kesehatan masyarakat adalah salah satu komponen kunci yang memengaruhi kualitas tenaga kerja dan, pada akhirnya, kemajuan suatu negara. Negara akan mencapai kesejahteraan jika status kesehatan penduduknya baik. Untuk meningkatkan kesehatan masyarakat, diperlukan kerja sama dan dukungan dari banyak pihak. Namun, pada dasarnya, setiap individu memiliki tanggung jawab terhadap kesehatannya sendiri. Hal ini karena kontribusi tiap individu akan sangat menentukan tinggi atau rendahnya status kesehatan masyarakat, yang menjadi fondasi bagi kesejahteraan (Sulistiarini, 2018).

Pada tahun 2010, Kementerian Kesehatan Republik Indonesia menciptakan sebuah indeks bernama Indeks Pembangunan Kesehatan Masyarakat (IPKM). Indeks ini terdiri dari 24 indikator yang datanya bersumber dari Riset Kesehatan Dasar (Riskesdas) 2007. IPKM berfungsi sebagai alat untuk mengukur pembangunan kesehatan masyarakat di tingkat regional maupun nasional. Model IPKM dengan 30 indikator dibuat pada tahun 2013. Tujuan dari pengembangan ini adalah untuk memperbaiki metode perhitungan dan melengkapi indikator penting yang belum disertakan. Menurut IPKM 2013, ada tujuh indeks kesehatan yang berbeda: balita, reproduksi, pelayanan kesehatan, perilaku kesehatan, penyakit tidak menular, penyakit menular, dan lingkungan. Sub-indeks ini memungkinkan

penghitungan dan pengukuran parsial dilakukan pada tingkat kabupaten, kota, dan provinsi (Indrawati dan Tjandrarini, 2018).

Nilai IPKM Provinsi Kalimantan Barat sedikit meningkat pada tahun 2018 menurut data (Riskesdas). Penyakit menular memiliki nilai tertinggi di sub indeks, dan pelayanan kesehatan memiliki nilai terendah. Sebagian besar sub-indeks mengalami peningkatan. Namun, ada kenaikan yang sangat signifikan pada sub-indeks kesehatan lingkungan dan kesehatan reproduksi. Ini mengindikasikan bahwa kondisi indikator-indikator yang membentuk kedua sub-indeks tersebut memang membaik. Sub indeks penyakit tidak menular, di sisi lain, mengalami penurunan yang signifikan, menunjukkan bahwa kondisi indikator penyusun sub indeks penyakit tidak menular memburuk (Kemenkes RI, 2019).

Berdasarkan penyebab penurunan IPKM tersebut, tujuan selanjutnya adalah menyederhanakan atau mengurangi jumlah variabel yang berkontribusi pada penurunan IPKM. Hal ini akan dilakukan dengan cara mempertahankan informasi sebanyak mungkin sekaligus menghilangkan korelasi antar variabel. *Principal Components Analysis* (PCA) adalah metode yang sesuai dengan tujuan tersebut. PCA (*Principal Component Analysis*) adalah kombinasi linear dari variabel-variabel awal. Secara geometris, ini bisa dibayangkan sebagai sistem koordinat baru yang didapat dari rotasi sistem koordinat sebelumnya. Metode ini sangat efektif saat ada banyak korelasi antar variabel dalam data. Dalam perhitungannya, PCA menghitung nilai eigen dan vektor eigen, yang berfungsi untuk menunjukkan penyebaran data dalam dataset (Johnson dan Wichern, 2007).

Karena perbedaan letak geografis dapat mempengaruhi aspek kesehatan, pemodelan persentase IPKM akan bergantung pada karakteristik daerah. Model *Geographically Weighted Regression* (GWR) adalah salah satu metode yang dapat digunakan untuk menganalisisnya (Stewart Fotheringham. A., Brunsdon. Chris, 2002).

Tujuan model *Geographically Weighted Regression* (GWR) adalah menjelajahi keragaman spasial melalui pembentukan model. Model GWR ini merupakan pengembangan dari model regresi linier. Perbedaannya, estimasi parameter pada model regresi linier berlaku secara global, sementara estimasi parameter model GWR bersifat lokal untuk setiap lokasi pengamatan (Purhadi dan Yasin, 2012). Salah satu jenis data yang digunakan dalam penelitian adalah data spasial, yang

berfokus pada posisi, objek, dan hubungan yang ada di Bumi ini. Metode regresi yang dapat digunakan untuk analisis data spasial adalah GWR (Kartika dan Kholijah, 2020).

Oleh karena itu, sebelum memulai pemodelan IPKM di Kalimantan Barat, analisis komponen utama yang dilakukan terhadap variabel-variabel yang mempengaruhi persentase IPKM harus dilakukan. Untuk menganalisis ini, penulis menggunakan metode GWRPCA, yang menggabungkan GWR dan PCA. Metode GWR dapat dipakai setelah uji heterogenitas spasial menunjukkan adanya perbedaan karakteristik antar wilayah. Jika ditemukan multikolinearitas lokal pada model GWR yaitu korelasi antar variabel independen di setiap lokasi pengamatan maka analisis komponen utama PCA dilakukan untuk menyederhanakan variabel-variabel independen yang berkorelasi tersebut. Selanjutnya, regresi dilakukan antara variabel dependen (y) dengan komponen utama terpilih, ditambah dengan penentuan bandwidth. Sebagai informasi, uji heterogenitas spasial dapat menggunakan uji *Breusch Pagan* (Sari et al., 2016).

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana model Indeks Pembangunan Kesehatan Masyarakat (IPKM) di Kalimantan Barat bisa dibentuk menggunakan metode *Geographically Weighted Regression* (GWR)?
2. Bagaimana pemodelan Indeks Pembangunan Kesehatan Masyarakat (IPKM) di Kalimantan Barat dilakukan dengan menerapkan metode *Geographically Weighted Regression Principal Component Analysis* (GWRPCA)?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dibuat, maka tujuan yang ingin dicapai melalui penelitian ini adalah sebagai berikut:

1. Untuk menentukan model Indeks Pembangunan Kesehatan Masyarakat (IPKM) di Kalimantan Barat menggunakan metode *Geographically Weighted Regression Principal Component Analysis* (GWRPCA).

2. Untuk menentukan bentuk terbaik dari GWRPCA pada data kesehatan di Kalimantan Barat.

1.4 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan dalam penelitian ini adalah data Riskesdas 2018 Kalimantan Barat.
2. Penentuan model dengan menggunakan metode GWRPCA.
3. Fungsi pembobot yang digunakan adalah fungsi kernel Gaussian.

1.5 Tinjauan Pustaka

Penggunaan *Geographically Weighted Regression Principal Component Analysis* (GWRPCA) telah digunakan oleh beberapa penelitian terdahulu untuk menentukan koefisien determinasi lokal dan global. Akan tetapi, penggunaan GWRPCA pada IPKM di Kalimantan Barat belum pernah dilakukan

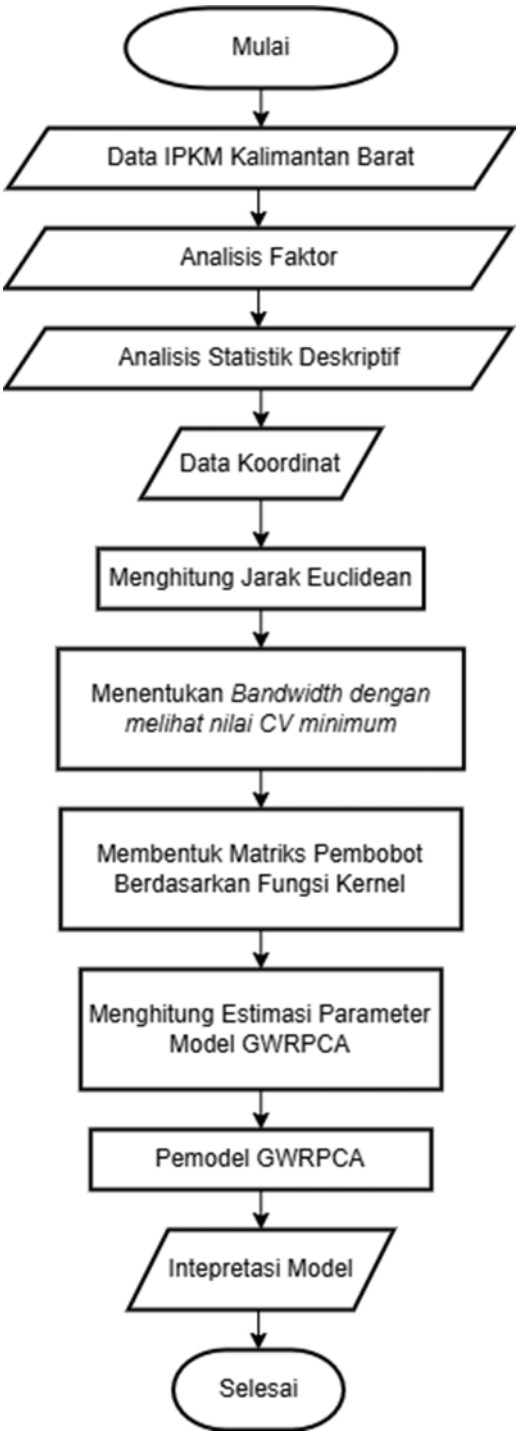
Sari, Yasin, dan Prahutama (2016) melakukan penelitian tentang pemodelan pendapatan asli daerah di Jawa Tengah menggunakan *Geographically Weighted Regression Principal Component Analysis* (GWRPCA). Hasil penelitian menunjukkan bahwa koefisien determinasi global GWRPCA untuk pemodelan Pendapatan Asli Daerah (PAD) di kabupaten/kota Provinsi Jawa Tengah adalah 0,9929. Ini berarti 99,29% dari total variasi PAD dapat dijelaskan oleh variabel-variabel seperti jumlah penduduk, retribusi daerah, belanja daerah, Produk Domestik Regional Bruto (PDRB) atas harga konstan, PDRB atas harga berlaku, dan pajak daerah. Sisa 0,0071% variasi dijelaskan oleh faktor-faktor lain di luar model.

Pratnyaningrum, Yasin, dan Suryowati (2015) melakukan penelitian tentang persentase balita gizi buruk di Jawa Tengah menggunakan *Geographically Weighted Regression Principal Component Analysis* (GWRPCA). Dengan metode GWRPCA, koefisien determinasi global untuk pemodelan persentase balita yang kekurangan nutrisi sebesar 0,6313309. Ini menunjukkan bahwa sebesar 63,13309% dari total persentase balita yang kekurangan nutrisi dijelaskan oleh variabel berikut: Sejumlah faktor yang memengaruhi status gizi balita antara lain jumlah bayi (0-6 bulan) yang menerima ASI eksklusif, jumlah bayi yang diimunisasi BCG, jumlah bayi yang

mendapatkan vitamin A, jumlah bayi yang menderita pneumonia, dan jumlah rumah tangga dengan akses air bersih. Perlu dicatat, 36,86691% dari total kasus balita kekurangan nutrisi disebabkan oleh faktor-faktor lain yang tidak termasuk dalam daftar tersebut. Di sisi lain, Kabupaten Sukoharjo memiliki koefisien determinasi lokal terendah, yaitu 0,03519539.

Luku dan Suryowati (2022) melakukan penelitian tentang pemodelan laju pertumbuhan PDRB di Provinsi Nusa Tenggara Timur menggunakan aplikasi metode *Geographically Weighted Regression Principal Component Analysis* (GWRPCA). hasil penelitian diperoleh nilai koefisien determinasi pada setiap wilayah lebih dari 50%. Sebagai contoh, di Kabupaten Rote Ndao, nilai koefisien determinasi adalah yang tertinggi, yaitu 99,63%. Ini berarti laju pertumbuhan PDRB di sana 99,63% dapat dijelaskan oleh variabel komponen utama yang baru. Sisa 0,37% dijelaskan oleh variabel lain di luar model yang diteliti.

1.6 Metodologi Penelitian



Gambar 1 *Flowchart* pemodelan indeks pembangunan kesehatan masyarakat dengan metode *Geographically Weighted Regression Principal Component Analysis* (GWRPCA)