PEMODELAN PADA DATA PRODUKSI KELAPA SAWIT DI KALIMANTAN BARAT MENGGUNAKAN GENERALIZED ADDITIVE MODEL

AHMAD FERNANDA H1091171043

SKRIPSI



PROGRAM STUDI STATISTIKA JURUSAN MATEMATIKA FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS TANJUNGPURA PONTIANAK 2023

PEMODELAN PADA DATA PRODUKSI KELAPA SAWIT DI KALIMANTAN BARAT MENGGUNAKAN GENERALIZED ADDITIVE MODEL

AHMAD FERNANDA H1091171043

SKRIPSI

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



PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2023

PEMODELAN PADA DATA PRODUKSI KELAPA SAWIT DI KALIMANTAN BARAT MENGGUNAKAN GENERALIZED ADDITIVE MODEL

Tanggung Jawab Yuridis, Material Pada

NIM.H1091171043

Disetujui Oleh

Pembimbing I,

Pembimbing II,

Naomi Nessyana Debataraja, S.Si, M.Si NIP. 198811232012122004

Hurry

Nur'ainul Miftabul Huda, S.Si., M.Si. NIP. 199411142020122014

Disahkan Oleh,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Tanjungpura Pontianak

NIP.197108022000031001

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET DAN TEKNOLOGI UNIVERSITAS TANJUNGPURA FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM PONTIANAK

TIM PENGUJI SKRIPSI

PEMODELAN PADA DATA PRODUKSI KELAPA SAWIT DIKALIMANTAN BARAT MENGGUNAKAN GENERALIZED ADDITIVE MODEL

AHMAD FERNANDA H1091171043

NAMA/NIP	TIM PENGUJI	GOLONGAN/ JABATAN	TANDA TANGAN
Naomi Nessyana Debataraja, S.Si., M.Si. NIP. 198811232012122004	Pemimpin sidang merangkap anggota penguji	III/c Lektor	Guar
Nur'ainul Miftahul Huda, S.Si., M.Si. NIP. 199411142020122014	Sekretaris sidang merangkap anggota penguji	III/b Asisten Ahli	Jay 12.
Ir. Dadan Kusnandar, M.Sc., Ph.D. NIP. 195907081987031014	Ketua penguji	IV/b Lektor Kepala	3K
Wirda Andani, S.Si., M.Si. NIP. 199411152022032016	Anggota penguji	III/b Tenaga Pengajar	Wh

Berdasarkan Surat Keputusan Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura Pontianak

Nomor

: 2590/UN22.8/PK.03.08/2023

Tanggal

: 21 Maret 2023

Tanggal Lulus

: 11 Mei 2023

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi. Sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain,kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Pontianak, M

Mei 2023

Ahmad Fernanda

PEMODELAN PADA DATA PRODUKSI KELAPA SAWIT DI KALIMANTAN BARAT MENGGUNAKAN GENERALIZED ADDITIVE MODEL

INTISARI

Kelapa sawit merupakan salah satu komoditas penting dan strategis peranannya dalam mendorong perekonomian rakyat di Kalimantan Barat. Banyak faktor yang mempengaruhi produksi kelapa sawit seperti luas lahan, produktivitas dan jumlah tenaga kerja pada perkebunan kelapa sawit. Tujuan penelitian ini untuk menerapkan Generalized Additive Models (GAM) pada data jumlah produksi kelapa sawit di Kalimantan Barat. Penelitian ini menggunakan metode GAM, yang merupakan perluasan dari Generalized linier Models (GLM) dengan menggantikan fungsi linier menjadi fungsi aditif dan mengganti prediktor non-linier dengan suatu fungsi penghalus. Penghalus yang digunakan pada permodelan adalah penghalus spline dan pendugaan parameter model dengan Peneralized Likelihood Maximation. Variabel respon penelitian ini adalah jumlah produksi dan variabel prediktor adalah luas lahan, produktivitas, dan tenaga kerja. Tenaga kerja memiliki pengaruh non-linier terhadap jumlah produksi. Sementara luas lahan dan produktivitas secara linier berpengaruh positif terhadap jumlah produksi. Nilai koefisien determinasi (R²) yang diperoleh dengan software R sebesar 0,957 yang berarti yang berarti kemampuan model dalam menggambarkan keragaman data sebesar 95,7%. Akan tetapi pada uji asumsi sisaan dapat dilihat model tidak memenuhi kehomogenan ragam sisaan yang dikatakan model tidak layak dalam mendeskripsikan data. Meskipun memiliki koefisien determinasi yang baik, model yang diperoleh tidak sesuai untuk mendeskripsikan data karena model yang diperoleh tidak efisien.

Kata kunci: Linier, Koefisien Determinasi, Model Aditif.

MODELING PALM OIL PRODUCTION DATA IN WEST KALIMANTAN USING A GENERALIZED ADDITIVE MODEL

ABSTRACT

Palm oil is one of the important and strategic commodities in boosting the economy of the people in West Kalimantan. Many factors affect oil palm production such as land area, productivity, and the amount of labor on oil palm plantations. The purpose of this study is to apply Generalized Additive Models (GAM) to data on the amount of palm oil production in West Kalimantan. This study uses the GAM method, which is an extension of Generalized Linear Models (GLM) by replacing linear functions into additive functions and replacing non-linear predictors with a smoothing function. The smoothing used in modeling is spline smoothing and estimation of model parameters with Peneralized Likelihood Maximation. The response variables of this study are the amount of production and the predictor variables are land area, productivity, and labor. Labor has a non-linear influence on the amount of production. Meanwhile, land area and productivity linearly have a positive effect on the amount of production. The value of the coefficient of determination (R2) obtained with software R is 0.957 which means that the model's ability to describe data diversity is 95.7%. However, in the test of residual assumptions, the model does not meet the homogeneity of the variety of leftovers that the model says is not feasible in describing the data. Despite having a good coefficient of determination, the obtained model is not suitable for describing data because the obtained model is inefficient.

Keywords: Linear, Coefficient of Determination, Additive Model.

PRAKATA

Alhamdulillah Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya serta shalawat dan salam tak lupa kita hanturkan kepada Nabi Muhammad SAW sehingga peneliti dapat menyelesaikan tugas akhir yang berjudul "Permodelan pada Data Produksi Kelapa Sawit di Kalimantan Barat Menggunakan Generalized Additive Model" dengan baik.

Penulisan skripsi ini disusun sebagai salah satu persyaratan untuk memperoleh gelar Sarjana Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura Pontianak. Selama penyusunan skripsi ini, penyusun telah banyak mendapat bimbingan dan bantuan dari berbagai pihak. Untuk itu pada kesempatan ini penyusun bermaksud menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

- 1. Allah SWT atas izin, rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi ini dengan lancar.
- 2. Kedua orang tua saya, serta Adik saya yang senantiasa memberikan do'a dan motivasi kepada penulis sehingga dapat terselesaikannya skripsi ini.
- 3. Ibu Naomi Nessyana Debataraja, M.Si., selaku dosen pembimbing pertama yang telah banyak memberikan masukan, arahan, dan motivasi untuk penulis dalam menyelesaikan skripsi ini.
- 4. Ibu Nur'ainul Miftahul Huda, M.Si., selaku dosen pembimbing kedua yang telah banyak memberikan masukan, arahan, dan motivasi untuk penulis dalam menyelesaikan skripsi ini.
- 5. Bapak Ir. Dadan Kusnandar, Ph.D., selaku dosen penguji I yang telah mengarahkan serta membimbing dan membagi ilmu-ilmunya dalam penulisan skripsi ini.
- 6. Ibu Wirda Andani, M.Si., selaku dosen penguji II yang telah mengarahkan serta membimbing dan membagi ilmu-ilmunya dalam penulisan skripsi ini.

7. Ahmad Mushawir, Paisal, Deny, Muhammad Septian, Wira Purnawan, Muhammad Asyorori, Valentinus Markus atas bantuan dan semangatnya

8. Teman teman-teman Statistika 2017 yang menemani saya selama berkuliah.

Penulis juga menyadari bahwa dalam penulisan skripsi ini masih terdapat banyak kekurangan yang masih jauh dari sempurna. Hal ini karena keterbatasan pengetahuan dan wawasan yang dimiliki penulis. Oleh karena itu, penulis masih membutukan setiap kritik dan saran yang dapat membangun dari pembaca untuk kebaikan kedepannya. Semoga skripsi ini bermanfaat bagi banyak pihak.

Pontianak, Januari 2023

Ahmad Fernanda

DAFTAR ISI

INTISARI	
ABSTRACT	i
PRAKATA	i
DAFTAR ISI	v
DAFTAR TABEL	vi
DAFTAR GAMBAR	vii
DAFTAR SIMBOL	ix
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	3
1.5 Tinjauan Pustaka	4
1.6 Metodologi Penelitian	5
BAB II LANDASAN TEORI	
2.1 Analisi Regresi	
2.2 Analisis Regresi Nonparametrik	8
2.3 Uji Multikolinearitas	8
2.4 Generalized Linier Models	9
2.5 Additive Models	10
BAB III METODE PENELITIAN	11
3.1 Generalized Additive Models (GAM)	11

	3.2 Penghalus <i>spline</i>	11
	3.3 Regresi Kubik siklik (Cyclic Cubic Regression)	12
	3.4 Penalized Iterative Reweighted Least Square (P-IRLS)	12
	3.5 Pemilihan Parameter Penghalus	123
	3.6 Uji Kelayakan Model	13
В	SAB IV HASIL DAN PEMBAHASAN	16
	4.1 Data	16
	4.2 Hubungan Variabel Prediktor dengan Variabel Respon	17
	4.3 Uji Linieritas	18
	4.4 Uji Asumsi Nonmultikolinieritas	19
	4.5 Validasi Kenormalan Distribusi Variabel Respon	20
	4.6 Uji Generalized Additive Model	20
	4.7 Uji Kelayakan Model	22
	4.7.1 Uji Asumsi Sisaan	22
	4.7.2 Uji Kelayakan Dimensi Basis (k)	23
	4.7.3 Uji Koefisien Determinasi (R ²)	23
	4.8 Interpretasi Hasil	24
В	SAB V KESIMPULAN DAN SARAN	25
	5.1 Kesimpulan	25
	5.2 Saran	25
D	OAFTAR PUSTAKA	26
r	A NADID A NI	20

DAFTAR TABEL

Tabel 4.1 Nilai uji F dengan pengujian Ramsey's RESET	19
Tabel 4 2 Hasil uji Nonmultikolinieritas	19
Tabel 4.3 Uji Generalized Additive Models dengan Software R	21
Tabel 4.4 Nilai-nilai parameter pemulus(k) yang dicoba	21

DAFTAR GAMBAR

Gambar 1. 1 Flowchart metode Generalized Additive Models	6
Gambar 4. 1 Statistik deskriftif	. 17
Gambar 4. 2 (a) Plot hubungan variabel Y dengan X1, (b) Plot hubungan variabe	:1 Y
dengan X2, (c) Plot hubungan variabel Y dengan X3.	. 17
Gambar 4. 3 Plot model Generalized Additive Models	. 25

DAFTAR SIMBOL

N : Jumlah data

 Y_i : Nilai ke-i variabel respon

 X_i : Nilai ke-i variabel prediktor

 ε : Galat

 $\boldsymbol{\theta}$: Fungsi parameter kanoik

β : Parameter

 f_i : Fungsi pemulusan

 \in_i : Nilai ke-i galat menyebar normal

g: Fungsi penghubung pemulusan

λ : Parameter pemulus

k : Banyaknya Knot

 \hat{f} : Fungsi penduga

BABI

PENDAHULUAN

1.1 Latar Belakang

Statistika adalah ilmu yang berkaitan dengan pengumpulan, penyajian, pengolahan, analisis data, dan interpretasi data menjadi informasi untuk membantu pengambilan keputusan yang efektif. Dalam Statistika itu sendiri terdiri dari berbagai macam metode dan analisis salah satunya analisis regresi. Analisis regresi merupakan salah satu bidang ilmu Statistika yang umumnya digunakan untuk menyelidiki pola hubungan fungsional antara variabel respon dan variabel prediktor. Terdapat tiga model pendekatan regresi, yaitu regresi parametrik, regresi nonparametrik, dan regresi semiparametrik (Budiantara, 2009). Salah satu bentuk model regresi nonparametrik adalah model aditif, jadi semua fungsi pada prediktor aditif merupakan fungsi pemulus (*smoothing*). Selain itu, model aditif juga dapat dibentuk menjadi model regresi semiparametrik. Penggunaan model aditif hanya digunakan pada data yang berdistribusi normal, sehingga dikembangkan *Generalized Additive Model* (GAM) untuk menangani kondisi tersebut, karenanya model aditif tergeneralisir lebih fleksibel data yang digunakan dibanding model aditif (Wigati, dkk, 2016)

GAM memiliki asumsi variabel respon berdistribusi keluarga ekponensial sama seperti pada *Generalized Linear Models* (GLM). Variabel respon pada GAM diasumsikan berdistribusi keluarga eksponensial dan hubungannya dengan variabel prediktor tidak harus linier. Pencocokan GAM menggunakan metode penghalusan berupa penghalus diagram pencar untuk mengestimasi fungsi penghalus, dan algoritma *Penalized Iteratively Re-Weighted Least Square* (P-IRLS) untuk mengestimasi model. Salah satu penghalus yang baik digunakan adalah penghalus *spline*, karena mempunyai banyak kelebihan dibandingkan penghalus lainnya, salah satunya yaitu dapat mengatasi pola data yang menunjukkan naik/turun yang tajam serta kurva yang dihasilkan relatif mulus. Oleh karenanya, banyak penelitian yang menggunakan model analisis tergeneralisir dengan penghalus *spline* untuk menganalisis data (Wood, 2006).

Pemilihan parameter penghalus yang baik digunakan dapat dilakukan dengan berbagai macam kriteria, salah satunya adalah menggunakan *Generalized Cross Validation* (GCV), *Cross Validation* (CV), *Akaike Information Criterion* (AIC). Metode GCV merupakan metode yang mendapatkan hasil maksimal pada parameter pemulus dari beberapa metode tersebut (Wood, 2006). Permasalahan jumlah produksi kelapa sawit juga tergolong sebagai kasus yang tidak diketahui hubungan antara variabel respon dan variabel prediktornya salah satunya adalah faktor-faktor yang mempengaruhi jumlah produksi kelapa sawit.

Kelapa sawit merupakan salah satu komoditas penting dan strategis di Kalimantan Barat karena peranannya yang cukup besar dalam mendorong perekonomian rakyat, terutama bagi petani perkebunan. Keberadaan perkebunan kelapa sawit di Kalimantan Barat melalui PT Perkebunan Nasional (PTPN XIII) dimulai sejak tahun 1980. Perusahaan ini diberikan tugas oleh pemerintah untuk mengembangkan lahan kelapa sawit dengan pola kerja sama terhadap masyarakat dengan pola Perkebunan Inti Rakyat (PIR). Kalimantan Barat merupakan provinsi dengan lahan perkebunan kelapa sawit terluas ketiga di Indonesia, dengan luas lahan 1,8 juta hektare. Perkebunan kelapa sawit terluas terdapat dari Provinsi Riau dengan luas 3,4 juta hektare dan Provinsi Sumatera Selatan 2,1 juta hektare (Kementerian Koordinator Bidang Perekonomian, 2019).

Faktor yang mempengaruhi hasil produksi tanaman kelapa sawit salah satunya adalah luas lahan. Menurut Suratiyah (2006) menyatakan bahwa semakin luas lahan yang dipakai maka semakin tinggi suatu produksi atau pendapatan perkesatuan luasnya. Beberapa tahun terakhir, luas perkebunan kelapa sawit di Indonesia terus bertambah. Pada tahun 2017, Kementerian Pertanian mencatat luas perkebunan kelapa sawit adalah 14.048.722 hektare. Tahun berikutnya naik menjadi 14.868.300 hektare (Direktorat Jenderal Perkebunan Kementerian Pertanian, 2022)

Selain luas lahan, faktor lain yang mempengaruhi produksi kelapa sawit adalah produktivitas dan tenaga kerja/petani. Produktivitas adalah faktor penting yang mempengaruhi kemampuan memproduksi kelapa sawit dalam sebuah perusahaan. Produktivitas merupakan kemampuan setiap orang, sistem, atau suatu perusahaan yang

dilakukan untuk menghasilkan barang atau jasa. Oleh karena itu, tenaga kerja memiliki peran besar dalam mengukur produktivitas. Tenaga kerja merupakan salah satu unsur penentu untuk kegiatan pertanian. Kelangkaan tenaga kerja dapat mengakibatkan mundurnya penanaman sehingga berpengaruh pada pertumbuhan tanaman, produktivitas dan kualitas produk (Suratiyah, 2006).

Salah satu faktor-faktor yang mempengaruhi jumlah produksi kelapa sawit antara lain adalah luas lahan, produktivitas terhadap perkebunan kelapa sawit dan jumlah tenaga kerja. Oleh karena itu, penelitian ini menerapkan *Generalized Additive Models* pada data jumlah produksi kelapa sawit di Provinsi Kalimantan Barat tahun 2019-2021 dan faktor-faktor yang mempengaruhinya. Dalam analisis ini digunakan *software R* untuk membantu menganalisis penelitian ini dengan menggunakan *package mgcv*.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang, maka rumusan masalah yang dibahas dalam penelitian ini adalah bagaimana penerapan metode GAM pada data produksi kelapa sawit di Kalimantan Barat.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini adalah untuk:

- Menentukan apakah GAM baik digunakan pada data produksi kelapa sawit di Kalimantan Barat.
- 2. Memprediksi faktor-faktor apa saja yang mempengaruhi jumlah produksi kelapa sawit di Kalimantan Barat menggunakan GAM.

1.4 Batasan Masalah

Berdasarkan dari tujuan penelitian ini terdapat beberapa batasan masalah yang diberikan yaitu:

- Data yang diambil adalah data produksi kelapa sawit di Kalimantan Barat pada tahun 2019 sampai 2021.
- 2. Penghalus *spline* adalah fungsi penghalus.
- 3. Pendugaan parameter model dengan Peneralized Likelihood Maximation

1.5 Tinjauan Pustaka

Nainggolan, Andriyana dan Bachrudin (2016) melakukan penelitian tentang penerapan *Generalized Additive Models* (GAM) pada rata-rata lama sekolah Provinsi Jawa Tengah. Tujuan dari penelitian ini adalah untuk menentukan kinerja *Generalized Additive Models* lebih optimal dibandingkan dengan *Generalized Linier Models* dan *Linier models*. Pada penelitian ini diperoleh MSE terkecil dihasilkan dengan menggunakan metode GAM yaitu sebesar 0,2856. R-square yang dihasilkan juga lebih besar dari kedua metode lainnya yaitu sebesar 0,8573.

Wigati, Tirta dan Anggraeni (2016) melakukan penelitian model aditif tergeneralisir dengan penghalus *spline*. Tujuan dari penelitian ini adalah untuk memahami bentuk dan asumsi model aditif tergeneralisir dengan penghalus spline. Model aditif tergeneralisir memiliki asumsi yang lebih fleksibel jika dibandingkan dengan model linier maupun model linier tergeneralisir, yaitu distribusi yang digunakan merupakan distribusi keluarga eksponensial dan hubungan antara variabel respon dengan variabel prediktor tidak harus linier. Dalam metode ini, model terbaik dapat diketahui melalui nilai AIC yang minimum. Selain itu, model dengan parameter penghalus optimal dapat dilihat melalui nilai GCV yang minimum.

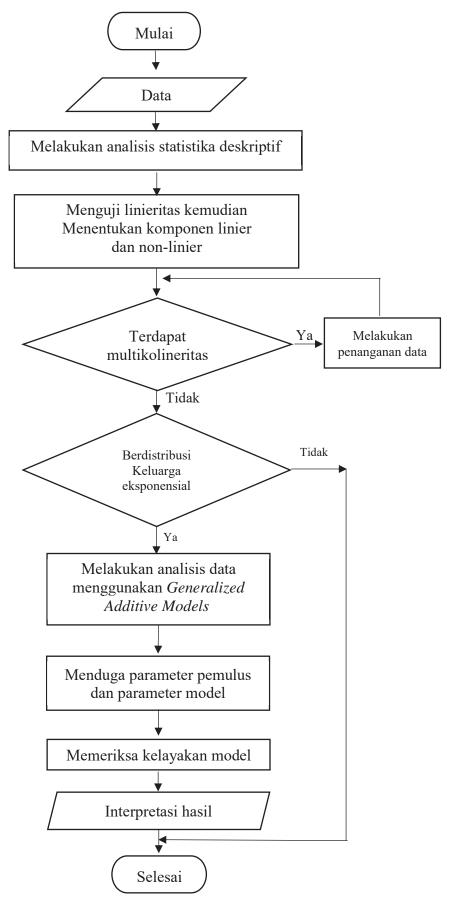
Safriana, Fajri dan Miftahudin (2017) melakukan penelitian tentang menggunakan GAM untuk menerapkan model pada data produksi padi di Indonesia. Tujuan penelitian ini adalah untuk menentukan pengaruh fiktur iklim terhadap produksi padi di Indonesia melalui model-model *Linear Models* (LM), GLM, dan GAM. Melalui penerapan basis spline kubik dikaji secara serempak temperatur udara, kelembaban, dan penyinaran matahari dan menunjukkan berpengaruh signifikan terhadap produksi padi di Indonesia (studi kasus: tahun 2013 dan 2014). Berdasarkan nilai AIC yang diperoleh, GAM sebagai model terbaik dengan nilai AIC 130,6588 dibandingkan model LM dan GLM yang memiliki nilai AIC 142,2529.

Valentina (2019) melakukan penelitian penerapan model aditif tergeneralisir pada studi kasus angka kematian bayi di Jawa Timur tahun 2015. Tujuan penelitian ini adalah banyaknya kasus persalinan yang menggunakan tenaga non medis memiliki pengaruh non-linear pada angka kematian bayi. Persentase perempuan yang menikah

dibawah 17 tahun dan lama pemberian ASI ekslusif secara linier berpengaruh positif terhadap peningkatan juga angka kematian bayi. Selain itu, kemampuan model dapat dengan baik dalam menjelaskan hubungan angka kematian bayi dan faktor-faktor yang mempengaruhinya. Model dikatakan baik dalam menggambarkan data jika mempunyai nilai koefisien determinasi lebih besar dari 0.5. Koefisien determinasi dari model aditif tergeneralisasi dalam peneltian sebesar 0.865. Artinya, kemampuan model dalam menjelasan keragaman data sebesar 86.5%.

1.6 Metodologi Penelitian

Data yang digunakan adalah data produksi kelapa sawit di Kalimantan Barat. Data dalam penelitian ini diperoleh dari website dinas perkebunan Kalimantan Barat https://disbunnak.kalbarprov.go.id/. Data tersebut adalah jumlah produksi sawit selama tiga tahun di kabupaten dan kota yang ada di Kalimantan Barat. Dengan variabel respon jumlah produksi kelapa sawit (Y) dan variabel prediktor adalah luas lahan perkebunan kelapa sawit (X_1) , produktivitas (X_2) dan jumlah tenaga kerja atau petani (X_3) . Setelah data diperoleh terlebih dahulu melakukan analisis deskriptif untuk menggambarkan karakteristik variabel respon dan prediktor. Selanjutnya, melihat pola hubungan antar variabel respon dan prediktor untuk menentukan komponen linier dan nonlinier. Menguji asumsi non-multikolinieritas dengan menghitung nilai Variance Inflation Factor (VIF). Kemudian langkah berikutnya adalah validasi kenormalan distribusi variabel respon (Y) dengan uji Kolmogorov Smirnov. Selanjutnya melakukan Generalized Additive Models dengan pemulus spline menggunakan metode Penalized Maximum Likelihood. Kemudian memeriksa kelayakan model berdasarkan R² dan pengujian asumsi residual dan terakhir melakukan interpratasi hasil. Metode penelitian disajikan dalam *flowchart* pada Gambar 1.1



Gambar 1. 1 Flowchart metode Generalized Additive Models