

BAB 1

PENDAHULUAN

1.1. Latar Belakang dan Identifikasi Masalah

Analisis regresi merupakan metode analisis data yang menggambarkan hubungan antara variabel respons dan variabel prediktor (Hosmer dan Lemeshow, 2000). Bentuk pola hubungan antara variabel prediktor dengan variabel respon dapat diidentifikasi berdasarkan informasi masa lalu atau dengan menggunakan *scatter plot* (Hardle, 1990).

Regresi dapat dilakukan dengan tiga pendekatan yaitu regresi parametrik, regresi semiparametrik, dan regresi nonparametrik. Pendekatan regresi parametrik membutuhkan asumsi-asumsi diantaranya bentuk kurva harus diketahui (mengikuti suatu pola tertentu misal linier, kuadratik, kubik), *error* berdistribusi normal, dan varians yang homogen. Suatu pendekatan regresi nonparametrik merupakan metode statistika yang digunakan untuk mengetahui hubungan antara variabel respon dengan variabel prediktor yang tidak diketahui bentuk fungsinya, hanya diasumsikan *smooth* (mulus) dalam arti termuat dalam suatu ruang fungsi tertentu. Sedangkan pendekatan regresi semiparametrik yaitu regresi yang memuat komponen parametrik dan komponen nonparametrik (Budiantara, 2009).

Apabila tidak ada informasi apapun mengenai bentuk kurva regresi apakah berbentuk linier, kuadratik atau yang lainnya, maka pendekatan yang digunakan adalah regresi nonparametrik. Pendekatan regresi nonparametrik tidak tergantung pada asumsi bentuk kurva tertentu, sehingga memberikan fleksibilitas

yang lebih tinggi, yang mana data diharapkan menyesuaikan sesuai bentuk estimasi kurva regresi tanpa dipengaruhi oleh subyektifitas peneliti. Dalam hal ini kurva regresi diasumsikan termuat dalam ruang fungsi tertentu (Budiantara, 2009). Keterbatasan informasi, bentuk fungsi, dan tidak jelasnya pola hubungan antara variabel respon dengan variabel prediktor merupakan pertimbangan sehingga digunakan pendekatan regresi nonparametrik.

Terdapat beberapa metode untuk memodelkan regresi dengan menggunakan pendekatan nonparametrik yaitu antara lain: *Kernel* (Hardle, 1990), *Spline* (Wahba, 1990; Budiantara, 2009), *K-Nearest Neighbor*(Hardle, 1990), *Histogram* (Green dan Silverman, 1994), Estimator *Deret Fourier* (Eubank,1988), *MARS*, *Deret Orthogonal*, *Wavelets*, *Neural Network*.

Diantara metode pendekatan tersebut, regresi nonparametrik dengan pendekatan *spline* merupakan metode yang sering digunakan. *Spline* merupakan salah satu model yang mempunyai model interpretasi statistik dan interpretasi visual sangat khusus dan sangat baik. *Spline* memiliki fleksibilitas yang tinggi. *Spline* juga memiliki kemampuan yang sangat baik dalam menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu(Budiantara, 2009).

Penelitian oleh Laome (2010) membandingkan model regresi nonparametrik *spline* dan *kernel*, menyatakan bahwa regresi *spline*, menghasilkan model yang lebih baik lebih baik dibandingkan dengan kernel. Penelitian (Mariati dan Budiantara, 2015) tentang pemodelan regresi *deret fourier* dan *spline tuncated* dalam regresi nonparametrik multivariabel pada kemiskinan di Papua, menyimpulkan bahwa metode *spline truncated* menghasilkan model lebih baik.

Hal ini dilihat dari nilai GCV yang minimum dan R^2 lebih besar dibandingkan dengan Deret Fourier. Perbandingan model regresi nonparametrik *Spline* dan Multivariate Adaptive Regression *Spline* (MARS) oleh Utami (2018), menyatakan bahwa model regresi *spline* terbaik dengan nilai MSE lebih kecil dibandingkan dengan menggunakan metode MARS.

Spline merupakan potongan-potongan polinomial yang memiliki sifat tersegmen dan kontinu. Pendekatan *spline* mempunyai suatu basis fungsi. Basis fungsi yang biasa digunakan antara lain *spline truncated* dan B-*spline*. *Spline Truncated* merupakan fungsi dimana terdapat perubahan pola perilaku kurva yang berbeda pada interval-interval yang berlainan. Berbagai macam kelebihan jika menggunakan *spline* antara lain: mempunyai interpretasi visual yang baik, bersifat fleksibel, serta mampu menangani karakter fungsi yang bersifat mulus. Selain itu, *spline* dapat mengatasi pola data dan dapat menggambarkan perubahan perilaku data yang berubah-ubah pada sub-sub interval tertentu. *Spline* juga mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik/turun yang tajam dengan bantuan titik-titik knot serta kurva yang dihasilkan relatif mulus (Eubank, 1999).

Titik knot adalah titik perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data (Budiantara, 2009). Dengan titik knot ini, *spline* dapat memberikan fleksibilitas yang lebih baik dari pada polinomial, sehingga memungkinkan untuk menyesuaikan diri secara efektif terhadap karakteristik lokal dari suatu fungsi atau data. Semakin optimal titik knot yang digunakan

dalam membuat model regresi maka akan semakin besar pula kebaikan model regresi yang didapatkan.

Ada tiga kriteria yang harus diperhatikan dalam membentuk model regresi *Spline*, antara lain: menentukan orde untuk model, banyaknya knot, dan lokasi penempatan knot, (Montoya, et all, 2014). Orde untuk model dapat ditentukan berdasarkan pola yang terjadi pada data, sementara banyaknya knot dan lokasi knot ditentukan berdasarkan perubahan pola data yang terjadi pada sub-sub interval tertentu. Terdapat beberapa metode untuk memilih titik knot optimal dalam regresi nonparametrik *Spline* antara lain metode *Cross Validation* (CV) (Craven dan Wahba, 1979), *Unbiased Risk* (UBR) (Wang, 1997), *Generalized Cross Validation* (GCV) (Wahba ,1990), dan *Generalized Maximum Likelihood* (GML) (Wahba ,1990).

Penelitian oleh Sari (2016) menunjukkan bahwa model *spline* dengan pemilihan titik knot optimal menggunakan metode GCV memberikan hasil yang lebih baik dibandingkan model *spline* dengan pemilihan titik knot optimal menggunakan metode UBR dengan pengujian pada 3 titik knot. Kesimpulan tersebut dengan memperhatikan kriteria kebaikan model, yaitu MSE yang relatif lebih kecil dan R^2_{adj} yang lebih besar. Penelitian lain (Hidayati, 2008) menyebutkan bahwa metode GCV adalah yang paling menghasilkan kurva yang *smooth*/mulus di bandingkan dengan metode GML, dengan titik knot optimal pada metode ini adalah $t=8$.

Penelitian Sifriani (2011) mengkaji estimator *spline* untuk mengestimasi fungsi variansi yang diasumsikan berdistribusi gamma dengan metode *Unbiased*

Risk (UBR) dan metode *Generalized Cross Validation* (GCV) untuk memilih parameter penghalus yang optimal pada fungsi variansi. Diselidiki pula *performance* (kebaikan) metode UBR dan metode GCV berdasarkan nilai MSE terkecil dengan menggunakan data simulasi. Berdasarkan hasil simulasi diperoleh untuk ukuran sampel 25 metode UBR lebih baik dibandingkan dengan metode GCV. Sedangkan untuk ukuran sampel 50 dan 100 diperoleh metode GCV lebih baik dibandingkan dengan metode UBR dalam pemilihan parameter penghalus untuk fungsi variansi. Penelitian lain oleh Nalim, Budiantara dan Fitriasari (2012) dan wang (1996) tentang model *spline* dengan *error* berkorelasi dengan menggunakan metode GML.

Saat ini analisis dengan menggunakan regresi nonparametrik *spline* banyak dimanfaatkan dalam berbagai sektor. Salah satunya di bidang kesehatan seperti penggunaan data Kematian Bayi. Pada penelitian oleh Megasari (2014) tentang pendekatan regresi semiparametrik *spline* linier untuk memodelkan angka kematian bayi di Jawa Timur, menunjukkan pola hubungan AKB dengan faktor yang diduga mempengaruhinya ada variabel yang menunjukkan pola hubungan linier dan variabel lainnya cenderung tidak membentuk pola tertentu. Hal ini mengidentifikasi bahwa terdapat kecenderungan pola nonparametrik karena fungsi kurva tidak diketahui.

Penelitian lainnya oleh Pratiwi (2017) tentang pemodelan *spline truncated* dalam regresi nonparametrik birespon. Penelitian ini bertujuan untuk mendeskripsikan AKB dan status gizi buruk balita serta faktor yang diduga mempengaruhinya. Pola hubungan yang terbentuk antara AKB dan angka gizi

buruk balita serta faktor yang mempengaruhinya menunjukkan pola hubungan yang tidak membentuk pola tertentu. Sehingga mengindikasikan bahwa terdapat komponen nonparametrik dimana fungsi dari kurva regresi tidak diketahui, maka pemodelan yang tepat adalah memodelkan dengan regresi nonparametrik birespon dengan estimator yang digunakan adalah *spline truncated*.

Salah satu permasalahan kesehatan yang menjadi indikator pembangunan kesehatan adalah Angka Kematian Bayi (AKB) yang masih tinggi secara nasional. Pada 2016, hasil riset Badan Pusat Statistik (BPS) mencatat bahwa angka kematian bayi (AKB) mencapai 25,5. Artinya, ada sekitar 25,5 kematian setiap 1.000 bayi yang lahir. Selama beberapa tahun terakhir, AKB Indonesia berangsur-angsur mengalami penurunan. Bahkan, perkembangan AKB di Indonesia cukup mengembirakan dalam waktu 20 tahun menunjukkan penurunan. Pasalnya, pada 1991 AKB pernah mencapai angka 68.

AKB di Indonesia masih termasuk tinggi dibandingkan dengan negara tetangga seperti Malaysia dan Singapura yang sudah di bawah 10 kematian per 1.000 kelahiran bayi. Kematian bayi merupakan salah satu indikator sensitif untuk mengetahui derajat kesehatan suatu negara dan bahkan untuk mengukur tingkat kemajuan suatu bangsa. Tingginya kematian bayi pada usia hingga satu tahun menunjukkan masih rendahnya kualitas sektor kesehatan di negara tersebut.

Angka Kematian Bayi (AKB) Indonesia tahun 2017 yaitu 24 per 1000 kelahiran hidup (KPPPA, 2018). Berdasarkan profil kesehatan Provinsi Kalimantan Timur tahun 2017, AKB di Kalimantan Timur yaitu 9 per 1000 kelahiran hidup. Sedangkan pada tahun 2018 jumlah kematian bayi di Provinsi

Kalimantan Timur tercatat sebanyak 671 orang dari 71.712 kelahiran hidup. Angka Kematian Bayi (AKB) menggambarkan tingkat permasalahan kesehatan masyarakat yang berkaitan dengan faktor penyebab kematian bayi, tingkat pelayanan antenatal, kemampuan pelayanan di fasilitas kesehatan dasar maupun di rumah sakit, status gizi ibu hamil, akses rujukan, tingkat keberhasilan program KIA-KB, kondisi lingkungan sosial dan ekonomi serta berbagai hal lainnya yang menjadi penyebab sekunder maupun tersier dari kematian bayi (Dinkes Provinsi Kalimantan Timur, 2018).

Berdasarkan penjelasan diatas, maka perlu dilakukan penelitian tentang perbandingan model regresi nonparametrik *spline* trunsated dengan menggunakan metode *Generalized Cross Validation* (GCV), *Unbiased Risk* (UBR) dan *Generalized Maximum Likelihood* (GML) yang didasarkan pada titik knot optimal, nilai R^2 adjusted terbesar dan *Mean Square Error* (MSE) terkecil dengan menggunakan data jumlah kematian bayi dan beberapa faktor yang mempengaruhinya untuk mendapatkan pemodelan prediksi jumlah kematian bayi di wilayah Provinsi Kalimantan Timur.

1.2. Kajian Masalah

Spline merupakan salah satu bagian dari analisis regresi, khususnya regresi nonparametrik yang digunakan ketika bentuk kurva regresi tidak diketahui. Regresi *spline* mempunyai interpretasi statistik dan interpretasi visual yang sangat khusus dan sangat baik. Selain itu *spline* memiliki fleksibilitas yang tinggi dan mampu menangani data/fungsi yang mulus (*smooth*). *Spline* juga memiliki

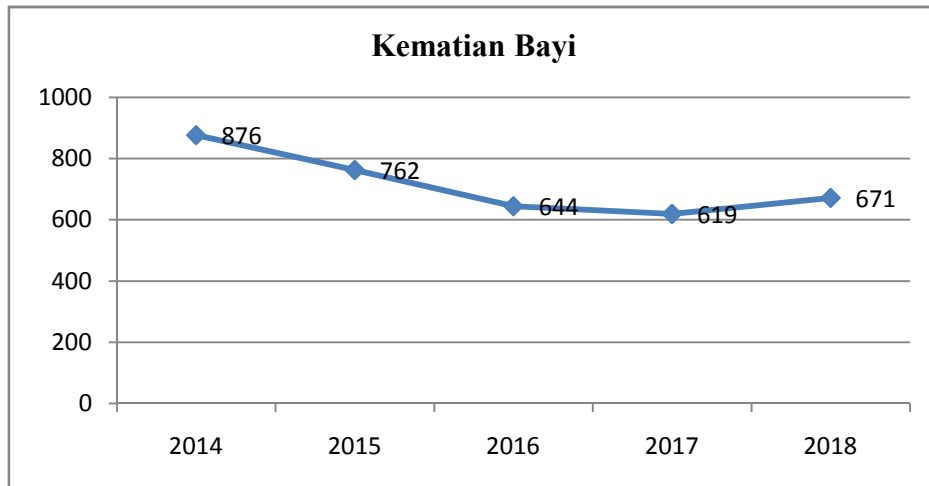
kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu karena *spline* merupakan salah satu jenis potongan polinomial, yaitu polinomial yang memiliki sifat tersegmen (Budiantara, 2009).

Kondisi data dengan bentuk kurva regresi yang tidak diketahui dapat ditemukan pada data kematian bayi. Seperti penelitian-penelitian terdahulu yang menyebutkan bahwa data kematian bayi dan variabel yang mempengaruhinya tidak membentuk pola tertentu sehingga menggunakan pendekatan regresi nonparametrik *spline* untuk memodelkan data kematian bayi (Megasari dan Budiantara, 2013)(Pratiwi, 2017).

Data Survei Demografi Indonesia (SDKI) tahun 2017 AKB di Indonesia mencapai 24 kematian per 1000 kelahiran hidup. Sedangkan berdasarkan Profil Kesehatan Kalimantan Timur, AKB tahun 2017 dan 2018 yaitu sebesar 9 per kematian per 1000 kelahiran hidup. Angka kematian bayi menjadi tantangan berat baik bagi pemerintah daerah maupun semua instansi terkait di Provinsi Kalimantan Timur dalam upaya menurunkan angka kematian bayi dan kesejahteraan penduduk di masa datang.

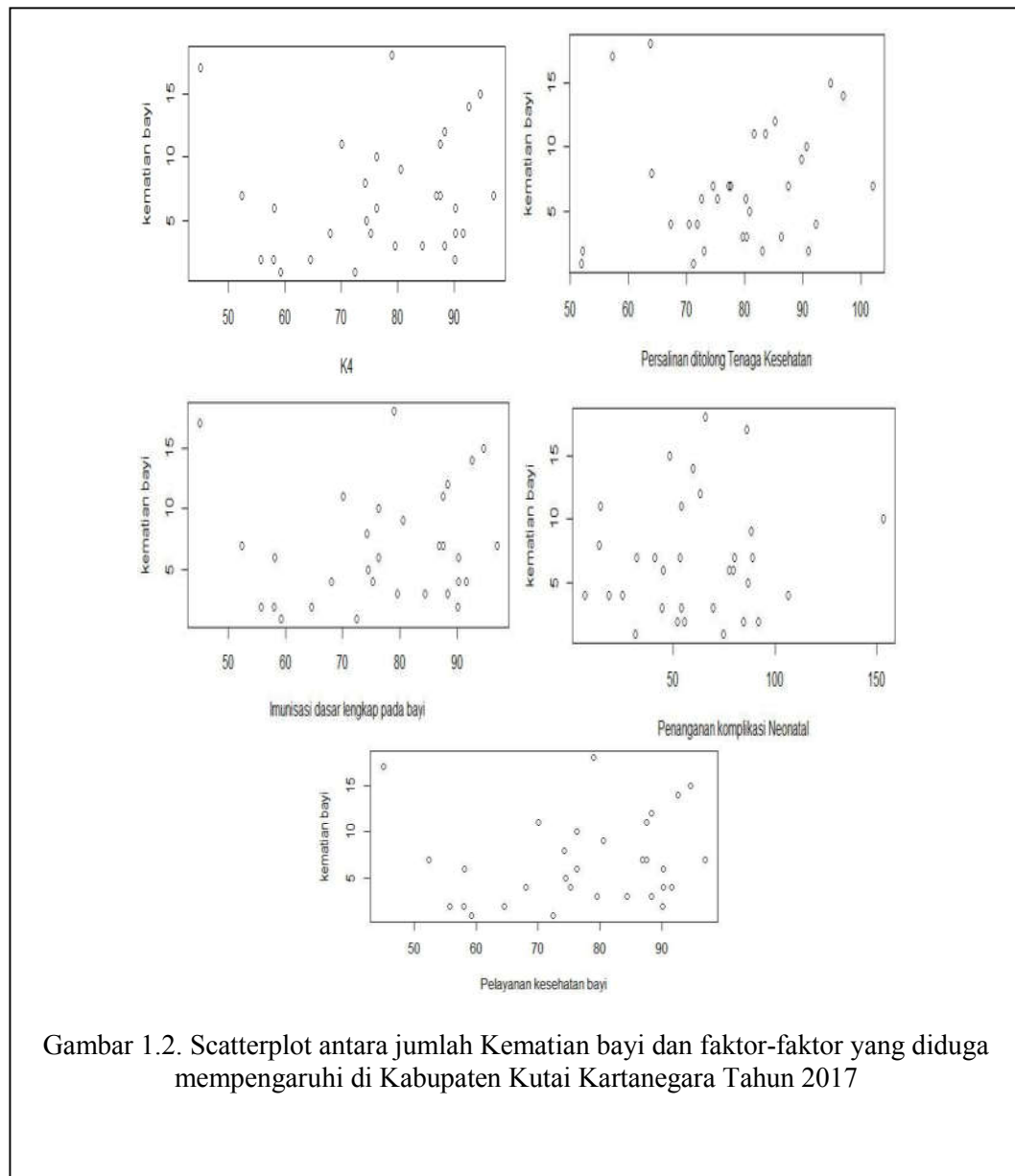
Kematian bayi di Provinsi Kalimantan Timur mengalami fluktuasi dari tahun 2014 - 2018, pada tahun 2014 kasus kematian bayi sebanyak 876 kematian dan menurun pada tahun 2015 menjadi 762 kasus kematian bayi. Pada tahun 2016 kasus kematian bayi sebesar 644 kasus kematian atau 9 per 1000 KH dan selanjutnya pada tahun 2017 jumlah kasus kematian menurun menjadi 619

kematian atau 9 per 1000 KH, sedangkan pada tahun 2018 terjadi peningkatan menjadi 671 kasus kematian.



Gambar 1.1. Jumlah Kematian Neonatal dan Bayi di Provinsi Kalimantan Timur Tahun 2014-2018

Pemilihan regresi nonparametrik pada penelitian ini berdasarkan studi awal yang dilakukan peneliti dengan melihat diagram pencar atau *scatter plot* dengan menggunakan data kematian bayi dan faktor yang diduga mempengaruhinya di Kabupaten Kutai Kartanegara Tahun 2017, dan diperoleh *scatter plot* sebagai berikut:



Gambar 1.2. Scatterplot antara jumlah Kematian bayi dan faktor-faktor yang diduga mempengaruhi di Kabupaten Kutai Kartanegara Tahun 2017

Gambar 1.2 menunjukkan pula hubungan antara kematian bayi dan faktor yang diduga mempengaruhi cenderung tidak membentuk pola tertentu. Sehingga mengindikasikan bahwa terdapat kecenderungan pola nonparametrik karena fungsi kurva regresi tidak diketahui. Hal ini akan berlaku sama pada lingkup yang lebih besar (provinsi). Selain itu ruang lingkup penelitian adalah puskesmas di Provinsi Kalimantan Timur, yang mana sumber variasi datanya adalah wilayah kerja

puskesmas. Oleh karena itu menyebabkan bentuk pola data akan acak dan tidak membentuk pola tertentu.

Berdasarkan konsep “*framework for child survival*” oleh Mosley and Chen yang mengemukakan bahwa terdapat 2 determinan yaitu determinan sosial-ekonomi atau eksogenous (seperti budaya, sosial, ekonomi, masyarakat dan faktor regional) dan determinan terdekat atau endogenous yang meliputi faktor ibu, faktor anak, faktor lingkungan, dan faktor pelayanan kesehatan. (Bappenas, 2009).

Penelitian lainnya oleh Pratiwi (2017) tentang pemodelan *spline truncated* dalam regresi nonparametrik birespon, menyimpulkan bahwa variabel yang berpengaruh terhadap AKB dan angka gizi buruk balita adalah persentase bayi yang diberi ASI dan persentase bayi BBLR.

Hasil penelitian oleh Rachmah, et al (2014) terkait pemodelan kematian bayi di Provinsi Jawa Timur diperoleh hasil bahwa yang mempengaruhi kematian bayi adalah persentase tenaga kesehatan, persentase persalinan oleh tenaga kesehatan, persentase ibu hamil yang melaksanakan K4, persentase rumah tangga ber-PHBS, persentase ibu hamil mendapatkan tablet Fe3, persentase wanita berstatus kawin di bawah 20 tahun, dan persentase KB aktif.

Penelitian oleh Lolong & Pangaribuan (2015), kunjungan K4 yang tidak terpenuhi beresiko 4,3 kali untuk melahirkan anak yang akan meninggal pada masa neonatal. Jika kunjungan K4 terpenuhi maka faktor risiko selama hamil dan pada saat melahirkan bisa ditatalaksana dengan baik sehingga dapat menurunkan risiko kematian neonatal dini. Selain itu, penelitian lain oleh Nugraha & Hayati

(2017) faktor yang berpengaruh terhadap jumlah kematian bayi antara lain adalah persalinan ditolong tenaga kesehatan dan penanganan komplikasi neonatal.

Kesehatan bayi harus selalu dipantau untuk memastikan kesehatan mereka selalu dalam kondisi optimal. Pelayanan kesehatan bayi termasuk salah satu dari beberapa indikator yang bisa menjadi ukuran keberhasilan upaya peningkatan kesehatan bayi dan balita. Sebagai salah satu kelompok yang menjadi sasaran program imunisasi, setiap bayi wajib mendapatkan imunisasi dasar Lengkap yang terdiri dari : 1 dosis BCG, 3 dosis DPT-HB dan atau DPT-HB-Hib, 4 dosis polio, dan 1 dosis campak. Imunisasi dasar lengkap yang diwajibkan tersebut, campak merupakan imunisasi yang mendapat perhatian lebih, komitmen Indonesia pada global untuk mempertahankan cakupan imunisasi campak sebesar 90% secara tinggi dan merata. Hal ini terkait dengan realita bahwa campak adalah salah satu penyebab utama kematian pada bayi. Sehingga pencegahan campak memiliki peran signifikan dalam penurunan angka kematian bayi (Dinkes Provinsi Kaltim,2018).

Berat badan lahir rendah (BBLR) merupakan faktor utama kematian perinatal dan neonatal. Semakin rendah berat lahir, probabilitas kelangsungan hidup neonatal juga semakin rendah. (Simbolon, 2012)

1.3. Rumusan Masalah

Berdasarkan latar belakang dan kajian masalah yang telah diuraikan, maka diperoleh rumusan masalah sebagai berikut, yaitu bagaimana perbandingan model regresi nonparametrik *Spline Truncated* dengan menggunakan metode *Generalized Cross Validation* (GCV), *Unbiased Risk* (UBR) dan *Generalized Maximum Likelihood* (GML) pada faktor yang mempengaruhi jumlah kematian bayi di Provinsi Kalimantan Timur tahun 2018.

1.4. Tujuan Penelitian

1.4.1. Tujuan Umum

Untuk mendapatkan model terbaik berdasarkan perbandingan model Regresi Nonparametrik *Spline Truncated* dengan menggunakan metode *Generalized Cross Validation* (GCV), *Unbiased Risk* (UBR) dan *Generalized Maximum Likelihood* (GML) pada faktor yang mempengaruhi jumlah kematian bayi di Provinsi Kalimantan Timur tahun 2018.

1.4.2. Tujuan Khusus

1. Mengidentifikasi titik knot optimal dan model Regresi Nonparametrik *Spline Truncated* dengan metode *Generalized Cross Validation* (GCV) pada faktor yang mempengaruhi jumlah kematian bayi di Provinsi Kalimantan Timur tahun 2018.
2. Mengidentifikasi titik knot optimal dan model Regresi Nonparametrik *Spline Truncated* dengan metode *Unbiased Risk* (UBR) pada faktor yang

mempengaruhi jumlah kematian bayi di Provinsi Kalimantan Timur tahun 2018.

3. Mengidentifikasi titik knot optimal dan model Regresi Nonparametrik *Spline Truncated* dengan metode *Generalized Maximum Likelihood* (GML) pada faktor yang mempengaruhi jumlah kematian bayi di Provinsi Kalimantan Timur tahun 2018.
4. Membandingkan model Regresi Nonparametrik *Spline Truncated* dari metode *Generalized Cross Validation* (GCV), metode *Unbiased Risk* (UBR) dan metode *Generalized Maximum Likelihood* (GML) pada faktor yang mempengaruhi jumlah kematian bayi di Provinsi Kalimantan Timur tahun 2018.

1.5. Manfaat Penelitian

1.5.1. Manfaat Keilmuan

1. Meningkatkan pengetahuan khususnya model Regresi Nonparametrik *Spline Truncated* dalam menentukan titik knot optimal pada data kematian bayi.
2. Analisis pemodelan Regresi Nonparametrik *Spline Truncated* merupakan salah satu upaya yang dapat memberikan alternatif penyusunan pemecahan masalah kesehatan yaitu angka kematian bayi di Provinsi Kalimantan Timur.

1.5.2. Manfaat Terapan

1. Sebagai bentuk pengembangan ilmu pengetahuan di bidang statistik dan kesehatan tentang analisis regresi nonparametrik *spline*.

2. Sebagai bantuan rujukan kepada Pemerintah, dalam hal ini Dinas Kesehatan Provinsi Kalimantan Timur dan sektor terkait lainnya agar dapat menggunakan informasi dari hasil penelitian ini untuk bahan pertimbangan dalam perencanaan program kesehatan khususnya terkait Angka Kematian Bayi.