

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Pada pemodelan regresi seringkali ditemui data dengan variabel tak bebas ( $Y$ ) yang tidak memiliki nilai pada sebagian observasinya atau dikenal dengan istilah data tersensor. Sering juga ditemui data dimana beberapa bagian nilai observasinya memiliki nilai tertentu yang bervariasi atau berada dibawah ambang batas (titik sensor) tertentu (data tersensor kiri), atau berada diatas ambang batas tertentu (data tersensor kanan). Penyensoran berganda (data tersensor interval) juga mungkin jika nilai dari variabel tak bebas ( $Y$ ) hanya tersedia untuk pengamatan dimana  $c < Y < d$ , dengan  $c$  dan  $d$  menjadi dua ambang batas. Dalam banyak kasus, ambang batas data tersensor kiri adalah nol, sedangkan ambang batas atas ditetapkan oleh peneliti sesuai dengan kasus yang diangkat [9].

Jika data tersensor dimodelkan menggunakan regresi linier maka kebanyakan observasi yang tidak memiliki nilai atau melewati batasan (limit) dianggap sebagai *outlier* dan tidak memenuhi uji asumsi klasik. Akibatnya akan sangat merugikan untuk membuang nilai-nilai variabel tak bebas yang tidak masuk dalam limit, ketika informasi tersebut tersedia [13]. Misal, data mengenai pengeluaran rumah tangga untuk barang-barang mewah dan bahan makanan yang

tergolong mahal. Jika data tersensor tersebut dimodelkan menggunakan regresi linier maka akan mengalami bias, atau memungkinkan terdapatnya masalah autokorelasi, heteroskedastisitas dan sebagainya. Oleh karena itu penting kiranya dilakukan pemodelan pada data tersensor dengan menggunakan metode pemodelan yang sesuai [14].

Metode regresi tobit (regresi tersensor) adalah salah satu metode pendekatan untuk memodelkan data tersensor. Penggunaan regresi tobit akan mengurangi efek bias karena data yang diasumsikan bernilai konstan dapat diolah secara bersamaan dengan data kontinu sehingga tidak akan kehilangan informasi yang berasal dari data diskrit. Dengan kata lain model regresi tobit dapat mengakomodasi semua observasi, baik yang bernilai nol maupun tidak nol [13].

Estimasi parameter regresi tobit menggunakan metode *Maximum Likelihood Estimation* (MLE) dengan memaksimalkan fungsi likelihood dapat menghasilkan estimasi yang konsisten dan efisien [13]. Selain itu metode MLE dengan basis bersyarat mean, merupakan estimasi yang diperoleh dengan menitikberatkan pada mean dari distribusi variabel tak bebas. Nilai mean menunjukkan ukuran pemusatan dari suatu distribusi sehingga hanya sedikit informasi yang diketahui dari keseluruhan distribusi. Oleh sebab itu pendekatan dengan metode ini hanya mampu menduga model dari fungsi bersyarat mean dan tidak mewakili keseluruhan data dari distribusi [17].

Selanjutnya dikembangkan metode regresi kuantil yang mana pada metode ini tidak membutuhkan asumsi residual dalam model dan estimasinya bersifat tegar terhadap variabel tak bebas yang mengandung pencilan (*outlier*)

[9]. Metode regresi kuantil pertama kali diperkenalkan oleh Koenker dan Bassett [25]. Pendekatan pada regresi kuantil adalah memisahkan atau membagi data menjadi kuantil-kuantil, dengan menduga fungsi kuantil bersyarat pada suatu sebaran data tersebut dan meminimumkan sisaan mutlak berbobot yang tidak simetris. Untuk mendapatkan estimasi parameter model regresi kuantil diperoleh dengan metode pemrograman linier yaitu metode simpleks [5] [27].

Metode regresi tobit digunakan untuk menduga parameter pada data tersensor. Namun karena estimasi yang diperoleh menitikberatkan pada mean dari distribusi variabel tak bebas yang mana estimasi parameter berbeda pada kondisi kuantil lebih rendah ataupun lebih tinggi, sehingga model regresi tobit menjadi kurang akurat untuk digunakan. Untuk mengatasi kekurangan tersebut berkembanglah model regresi tobit kuantil yang dapat mengatasi permasalahan distribusi error yang bersifat heteroskedastisitas, tidak normal dan data tidak simetris [36]. Model regresi tobit kuantil diperkenalkan oleh Powell (1986), sehingga dikenal sebagai estimasi Powell [36].

Estimasi regresi tobit kuantil bisa diperoleh menggunakan solusi minimasi metode pemrograman linier yaitu metode *Iterative Linear Programming Algorithm* (ILPA) yang melibatkan *Barrodale-Robert Algorithm* (BRA) dan algoritma BRCENS sebagai adaptasi dari algoritma BRA. Namun demikian seluruh algoritma tersebut memberikan performa yang kurang baik pada kondisi proporsi data tersensor yang besar [5] [12]. Dalam mengatasi kekurangan pada model regresi tobit kuantil tersebut, kemudian dihibrid dengan metode Bayesian. Pendekatan Bayesian dalam menduga parameter pada regresi tobit

kuantil mendapat perhatian dalam literatur karena cenderung menghasilkan dugaan model yang lebih baik [31] [55] [54] [29] [11]. Bayes memperkenalkan suatu metode untuk mengestimasi parameter yang memanfaatkan informasi awal yang disebut distribusi awal (prior). Metode ini dikenal dengan metode Bayesian. Distribusi prior ini dapat berasal dari data penelitian sebelumnya atau berdasarkan intuisi seorang peneliti. Informasi prior dari sebaran parameter tersebut kemudian digabungkan dengan informasi dari data yang didapat dari pengambilan sampel atau yang disebut juga dengan fungsi likelihood sehingga didapat distribusi posterior dari parameter. Rataan dan varian dari distribusi posterior ini atau posterior mean dan posterior varian yang kemudian menjadi penduga bagi parameter regresi dengan metode Bayesian [51]. Metode Bayesian menggunakan pendekatan algoritma MCMC (*Markov Chain Monte Carlo*) untuk menduga distribusi posterior dari parameter yang memiliki formulasi kompleks yang susah diestimasi posterior mean dan posterior varian secara analitik.

Terdapat beberapa penelitian terdahulu terkait metode regresi tobit kuantil Bayesian yang pernah dilakukan. Cintiani [7] telah membahas pemodelan regresi kuantil tersensor pada pengeluaran rumah tangga untuk konsumsi rokok. Lusiana [31] telah membahas metode regresi tobit kuantil Bayesian pada pengeluaran rumah tangga untuk konsumsi susu dengan menggunakan data yang diperoleh dari hasil Survei Sosial Ekonomi Rumah Tangga Indonesia (SUSETI) tahun 2011. Selanjutnya Yue dan Hong [55] memodelkan data pengeluaran rumah tangga untuk kesehatan menggunakan regresi tobit kuantil

Bayesian. Yu dan Stander [54] melakukan perbandingan estimasi parameter model regresi tobit kuantil Bayesian dengan estimasi Powell dengan metode ILPA menggunakan teknik simulasi.

Penyakit COVID-19 (*Coronavirus Disease 2019*) merupakan penyakit menular sindrom pernafasan akut yang disebabkan oleh SARS-CoV-2 yang termasuk keluarga besar Coronavirus penyebab SARS pada tahun 2003. Penyakit ini awalnya mewabah di Wuhan, China pada Desember 2019 dan kini telah menyebar ke seluruh dunia [56]. Organisasi Kesehatan Dunia (WHO) telah menetapkan COVID-19 sebagai pandemi global sejak 11 Maret 2020 karena penyebarannya yang telah meluas ke seluruh dunia termasuk Indonesia [47]. Kasus COVID-19 di Indonesia pertama kali dikonfirmasi pada awal bulan Maret 2020, seiring berjalannya waktu COVID-19 terus menyebar luas hingga ke Sumatera Barat.

Menurut Data Dinas Kesehatan Provinsi Sumatera Barat per 30 November 2020, tercatat sebanyak 20.036 kasus positif terinfeksi COVID-19 di Sumatera Barat dengan rincian 378 orang dirawat di berbagai rumah sakit, 1.995 orang isolasi mandiri, 80 orang isolasi Provinsi, 168 orang isolasi Kab/Kota, 431 orang meninggal dunia dan 16.984 orang sembuh [35]. Tingginya angka kasus positif terinfeksi COVID-19 di Sumatera Barat mengakibatkan banyaknya pasien COVID-19 yang menjalani isolasi dan perawatan di rumah maupun di rumah sakit. Lama rawat inap pasien berbeda-beda untuk setiap pasien karena dipicu oleh beberapa faktor [35]. Oleh karena itu perlu dilakukan kajian tentang faktor-faktor yang mempengaruhi lama rawat inap pasien COVID-19 di rumah

sakit. Pemodelan lama rawat inap ini penting bagi pihak rumah sakit untuk penyediaan jumlah tempat tidur, penyediaan layanan medis, staff dan peralatan medis lainnya di rumah sakit.

Penelitian terkait lama rawat pasien terinfeksi COVID-19 sudah dilakukan oleh banyak peneliti diantaranya oleh Thiruvengadam dkk [43] yang menganalisis tentang faktor yang mempengaruhi lama rawat inap pasien COVID-19 dengan model hazard cox-proportional di rumah sakit perawatan tersier India Selatan. Guo dkk [15] menganalisis faktor lama rawat inap pada pasien COVID-19 menggunakan studi retrospektif. Wu dkk [48] menganalisis faktor-faktor yang mempengaruhi lama rawat pasien COVID-19 yang tidak parah di rumah sakit penampungan Fangcang, Cina.

Permasalahan yang diangkat pada penelitian ini adalah bagaimana memodelkan data tersensor yang terdapat pada data lama rawat inap pasien COVID-19 dan mengidentifikasi faktor-faktor yang mempengaruhi lama rawat inap pasien COVID-19 menggunakan metode regresi tobit kuantil dan metode regresi tobit kuantil Bayesian. Penelitian ini juga akan mengevaluasi kebaikan model yang dihasilkan oleh metode regresi tobit kuantil dan regresi tobit kuantil Bayesian menggunakan metode evaluasi kebaikan model RMSE dan  $PseudoR^2$ .

## 1.2 Perumusan Masalah

Adapun rumusan masalah penelitian ini antara lain:

1. Bagaimana mengidentifikasi faktor-faktor yang mempengaruhi lama rawat inap pasien COVID-19 menggunakan metode regresi tobit kuantil dan

metode regresi tobit kuantil Bayesian ?

2. Bagaimana menentukan evaluasi kebaikan model yang dihasilkan?

### 1.3 Tujuan Penelitian

Adapun tujuan penelitian ini adalah:

1. Mengetahui faktor-faktor yang mempengaruhi lama rawat inap pasien COVID-19 menggunakan metode regresi tobit kuantil dan metode regresi tobit kuantil Bayesian.
2. Mengetahui kebaikan model yang dihasilkan menggunakan metode RMSE dan  $PseudoR^2$ .

### 1.4 Sistematika Penulisan

Penelitian ini disusun dengan sistematika yang terdiri dari lima bab, yaitu BAB I Pendahuluan yang berisi latar belakang, perumusan masalah, tujuan penelitian dan sistematika penulisan. BAB II Landasan Teori berisi teori-teori yang terkait dalam pembahasan dan mendukung masalah yang dibahas dan BAB III Metode Penelitian yang memaparkan tentang cara penyelesaian masalah yang telah dirumuskan. Pada BAB IV Hasil dan Pembahasan, akan dibahas mengenai gambaran umum faktor-faktor yang mempengaruhi lama rawat inap pasien COVID-19, pemodelan lama rawat inap pasien COVID-19 menggunakan metode tobit kuantil dan metode regresi tobit kuantil Bayesian dan mengevaluasi kebaikan model menggunakan metode RMSE.

BAB V Penutup, memberikan kesimpulan berdasarkan hasil yang diperoleh pada pembahasan dan juga disampaikan saran yang menjadi pedoman untuk penelitian selanjutnya.

