

BAB I

PENDAHULUAN

1.1 Latar Belakang

Regresi merupakan suatu metode dalam statistika untuk mendapatkan suatu fungsi yang dapat memprediksi variabel respon/dependen terhadap satu atau lebih variabel independennya. Analisis regresi meliputi tiga aspek utama, yaitu pemilihan variabel-variabel bebas, transformasi, dan estimasi parameter dalam model regresi. Pemilihan model regresi terbaik dapat dinyatakan sebagai optimasi masalah dengan tujuan memilih variabel-variabel independennya untuk memaksimalkan kriteria-kriteria yang dibutuhkan.

Metode yang biasa digunakan dalam pemilihan model terbaik yaitu menggunakan *backward regression*. *Backward regression* merupakan metode pengoptimalan model regresi yang menyeleksi variabel bebas pada model satu per satu hingga didapatkan nilai *Akaike Information Criterion* (AIC) terkecil. Namun metode tersebut memiliki kelemahan jika terdapat banyak variabel bebas dalam model. Proses pemilihan variabel satu per satu secara berurutan mempersempit kemungkinan pemilihan model optimal yang dapat ditemukan jika memperluas ruang sampel. Pengerjaan seperti ini disebut proses pencarian lokal (*local search process*) dan membuat optimasi kurang maksimal.

Minerva dan Paterlini (2010) menyodorkan alternatif dalam pengoptimalan model regresi menggunakan algoritme genetika. Algoritme genetika merupakan bagian dari algoritme evolusi yang mengandalkan operator-operator seperti seleksi mutasi, dan pindah silang atau *crossover* (Sivanandam&Deepa:

2008). Algoritme genetika memiliki beberapa kelebihan. Pertama, algoritme genetika merupakan algoritme stokastik. Pada algoritme genetika, pemilihan faktor acak (*random*) merupakan hal yang utama. Proses seleksi dan reproduksi dilakukan secara acak. Kedua, algoritme genetika selalu menawarkan solusi dalam bentuk populasi, yang berarti solusi yang dihasilkan dapat lebih dari satu pada setiap iterasinya. Ketiga, algoritme genetika mempunyai sifat *robust* yang berarti tidak ada persyaratan khusus untuk menggunakan algoritme tersebut.

Berdasarkan hasil penelitian Minvera dan Paterlini, optimasi model regresi menggunakan algoritme genetika dapat menghasilkan transformasi terbaik terhadap variabel-variabel independennya karena algoritme tersebut mengeksplorasi populasi atau kumpulan solusi yang saling berkompetisi satu sama lain dan bergabung satu sama lain. Proses ini juga memungkinkan tercapainya nilai global maksimum.

Penelitian serupa dilakukan oleh Fraticasari, Ratnawati, dan Wihandika (2018) yang mengoptimasi model regresi linear berganda pada prediksi jumlah kecelakaan sepeda motor dengan algoritme genetika. Dari penelitian tersebut, hasil eror dari model regresi linear berganda lebih besar dibanding eror pada menggunakan algoritme genetika. Budianto dkk. (2017) juga menambahkan kasus pencilan dalam data pengamatan untuk regresi linear berganda. Kasus tersebut diatasi menggunakan algoritme genetika dengan fungsi *fitness* berupa kriteria informasi sehingga data pencilan dapat terdeteksi dan nilai koefisien determinasi model lebih besar. Namun kedua penelitian tersebut menggunakan model linear yang mengasumsikan variabel responnya kontinu serta sebaran data normal. Sedangkan tidak semua penelitian menggunakan data kontinu dan memiliki sebaran normal.

Bentuk pengembangan dari model regresi klasik yang mampu menyelesa-

ikan permasalahan di atas, khususnya dalam menangani variabel respon yang tidak normal adalah dengan menggunakan model linier tergeneralisir atau disebut *Generalized Linear Model* (GLM). Salah satu bentuk GLM adalah model regresi Poisson. Model regresi Poisson mengasumsikan data variabel respon berdistribusi Poisson. Dalam penelitian Simonoff (2003), regresi Poisson dibangun atas data jumlah kematian akibat tornado sebagai variabel respon. Sedangkan tahun serta klasifikasi tingkat bahaya tornado menjadi variabel-variabel bebas pada model.

Pada pemodelan regresi Poisson, sering muncul masalah yang kerap muncul dalam pengolahan data berdistribusi Poisson. Distribusi Poisson mengasumsikan data memenuhi kondisi equidispersi, dimana rata-rata dan variansi bernilai sama, yaitu μ . Jika kondisi ini tidak dipenuhi, maka hasil pengolahan data menjadi tidak valid, sehingga dibutuhkan solusi untuk mengatasi kasus penyimpangan equidispersi. Model *Generalized Poisson Regression* diajukan oleh Felix Famoye (1993) untuk mengatasi kasus penyimpangan equidispersi, yaitu dengan menambahkan variabel dispersi θ pada model regresi Poisson sehingga dapat mengukur seberapa jauh penyimpangan yang terjadi pada model.

Sebelumnya, telah dilakukan juga penelitian terhadap regresi Poisson menggunakan algoritme genetika dalam jurnal Peterson (2014). Peterson menggunakan algoritme genetika sebagai metode seleksi variabel pada model regresi Poisson. Nilai *fitness* yang digunakan pada optimasi regresi Poisson menggunakan algoritme genetika adalah nilai devians pada model. Sedangkan model yang dihasilkan pada proses algoritme genetika dibandingkan nilai AIC-nya dengan model regresi Poisson menggunakan metode Newton-Raphson. Hasil dari penelitian tersebut menyatakan bahwa nilai AIC pada regresi Poisson menggunakan algoritme genetika lebih kecil dari AIC pada regresi Poisson

menggunakan Newton-Raphson.

Berdasarkan penjabaran di atas, penulis tertarik untuk melakukan penelitian mengenai *Optimasi Model Generalized Poisson Regression Menggunakan Algoritme Genetika*. Dalam penelitian ini, model *Generalized Poisson Regression* akan dioptimalkan menggunakan algoritme genetika. Kemudian, dilakukan perbandingan hasil optimasi model regresi *Generalized Poisson Regression* dengan *generalized Poisson regression* menggunakan algoritme genetika untuk mengetahui hasil yang paling baik di antara keduanya.

1.2 Perumusan Masalah

Berdasarkan latar belakang permasalahan di atas, maka dapat dirumuskan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana mengestimasi parameter pada model *Generalized Poisson Regression*?
2. Bagaimana mengoptimasi model *Generalized Poisson Regression* menggunakan algoritme genetika?
3. Bagaimana hasil optimasi model *Generalized Poisson Regression* menggunakan algoritme genetika dibanding model *generalized Poisson regression*?

1.3 Tujuan Masalah

Tujuan dari penulisan ini dapat dirumuskan sebagai berikut:

1. Mengetahui cara estimasi parameter pada model *Generalized Poisson Regression*.

2. Mengetahui cara mengoptimasi model *Generalized Poisson Regression* menggunakan algoritme genetika.
3. Mengetahui perbedaan hasil optimasi model *Generalized Poisson Regression* dengan *generalized Poisson regression* menggunakan algoritme genetika.

1.4 Pembatasan Masalah

1. Data variabel respon yang digunakan adalah data diskret.
2. Data variabel bebas pada model regresi adalah data diskret dan kontinu non-kategorikal.
3. Model *Generalized Poisson Regression* digunakan untuk mengatasi kasus penyimpangan equidispersi.
4. Data variabel respon yang digunakan tidak mengikuti distribusi Poisson.
5. Estimasi parameter model regresi menggunakan metode Newton-Raphson.
6. Algoritme genetika pada *generalized Poisson regression* dilakukan untuk seleksi variabel bebas pada model.
7. Perhitungan nilai *fitness* pada algoritme genetika menggunakan nilai devians.
8. Kriteria stop pada proses algoritme genetika adalah ukuran maksimum generasi (*maxgen*).
9. Data yang digunakan untuk aplikasi data merupakan data sekunder mengenai kasus AIDS di Jawa Tengah tahun 2019.

1.5 Manfaat Penelitian

1. Bagi penulis sendiri, penelitian ini merupakan penerapan teori yang telah dipelajari dengan keadaan yang sebenarnya.
2. Bagi pembaca, penelitian ini dapat digunakan sebagai bahan referensi dan pembandingan studi maupun penelitian yang terkait dengan riset ini.
3. Bagi universitas, dapat dijadikan salah satu referensi karya tulis khususnya untuk Program Studi Matematika.

1.6 Metode Penelitian

Skripsi ini berupa kajian pustaka dalam bidang statistika dan algoritme pemrograman dengan mengumpulkan berbagai literatur dalam bentuk jurnal dan buku yang mendukung penelitian ini. Acuan utama yang digunakan dalam penulisan ialah buku *Introduction to Genetic Algorithms* (Sivanandam & Deepa: 2008) dan *Restricted Generalized Poisson Regression Model* (Famoye: 1993). Metode yang digunakan adalah metode regresi *Generalized Poisson Regression* menggunakan algoritme genetika untuk mendapatkan model terbaik.