

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi adalah salah satu metode untuk menentukan hubungan sebab akibat antara variabel respon dengan variabel prediktor. Model matematis yang menyatakan hubungan antara kedua variabel tersebut disebut dengan persamaan regresi [7]. Hubungan antara variabel respon dan variabel prediktor pada regresi dimanfaatkan untuk membuat prediksi pada berbagai permasalahan penelitian yang ada di lapangan. Variabel respon pada model regresi ada yang bersifat kontinu dan ada juga bersifat kategorik.

Model regresi untuk variabel respon bersifat kategorik adalah regresi logistik [9]. Pada regresi logistik, peneliti memprediksi variabel respon yang bersifat dikotomis atau polikotomis. Regresi logistik terdiri dari dua jenis yaitu regresi logistik biner dan regresi logistik ordinal/multinomial.

Regresi logistik biner merupakan suatu model yang digunakan untuk menentukan hubungan antara variabel respon yang bersifat biner (dikotomis) dengan variabel prediktor. Variabel respon yang bersifat biner memiliki dua kemungkinan hasil. Contoh variabel respon yang dimaksud adalah kesuksesan (sukses-gagal), kesetujuan (setuju-tidak setuju), keinginan membeli (ya-tidak), terpilih atau tidak terpilih dan masih banyak lagi. Oleh karena itu, variabel

respon mengikuti distribusi binomial. Pada regresi logistik terdapat beberapa asumsi, salah satunya yaitu asumsi non multikolinieritas.

Multikolinieritas merupakan salah satu pelanggaran asumsi klasik pada analisis regresi linier berganda, hal ini disebabkan karena adanya korelasi atau hubungan linier antara dua atau lebih variabel bebas dalam sebuah model regresi berganda. Ini merupakan sebuah masalah dalam regresi logistik karena dapat berpengaruh pada tingkat akurasi prediksi model dan menyebabkan kesalahan dalam pengambilan keputusan. Model yang baik adalah model yang bebas dari multikolinieritas [8].

Ada beberapa metode yang dapat mengatasi masalah multikolinieritas pada regresi linear berganda. Diantaranya yaitu Metode LASSO (*Least Absolute Shrinkage and Selection Operator*), PCA (*Principal Component Analysis*), metode *regresi ridge* dan metode *regresi bayesian* [18]. Metode *Least Absolute Shrinkage and Selection Operator* (LASSO) dapat menyusutkan koefisien regresi pada variabel prediktor menjadi mendekati nol atau bahkan sama dengan nol sehingga dapat melakukan seleksi variabel sekaligus mengatasi multikolinieritas[20].

Pada penelitian ini akan digunakan Metode LASSO untuk mengestimasi koefisien regresi logistik biner pada data tingkat pengangguran terbuka. Pengangguran terbuka merupakan salah satu masalah dalam pembangunan ekonomi. Pengangguran adalah masalah yang sangat kompleks karena mempengaruhi sekaligus dipengaruhi oleh banyak faktor yang saling berinteraksi. Jika pengangguran tidak segera diatasi maka dapat menimbulkan masalah lain,

seperti kerawanan sosial dan berpotensi mengakibatkan kemiskinan.

Terjadinya pengangguran di suatu daerah dapat disebabkan oleh beberapa hal. Salah satunya dikarenakan jumlah lapangan pekerjaan di suatu wilayah tertentu tidak dapat mencukupi jumlah angkatan kerja atau jumlah permintaan akan lapangan pekerjaan dengan penawaran lapangan kerja tidak seimbang. Hal tersebut mengakibatkan jumlah pertumbuhan tenaga kerja yang melebihi jumlah kesempatan kerja meningkat[13].

Pulau Sumatera merupakan salah satu dari lima pulau terbesar yang terdapat di Indonesia dan terdiri dari sepuluh provinsi. Dari sepuluh provinsi tersebut terdapat empat provinsi yang memiliki jumlah kabupaten/kota cukup banyak dibanding provinsi lain. Empat provinsi tersebut yaitu Provinsi Sumatera Barat, Provinsi Sumatera Selatan, Provinsi Lampung dan Provinsi Riau. Sama halnya dengan negara Indonesia dan provinsi lainnya, empat provinsi tersebut juga tidak lepas dari permasalahan pengangguran. Oleh karena itu, analisis regresi logistik biner dengan Metode LASSO pada penelitian ini diaplikasikan pada data tingkat pengangguran terbuka empat provinsi di Pulau Sumatera.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang diuraikan sebelumnya, masalah yang dibahas pada penelitian ini adalah apa saja faktor-faktor yang berpengaruh signifikan terhadap tingkat pengangguran terbuka empat provinsi di Pulau Sumatera.

1.3 Batasan Masalah

Dalam penelitian ini, variabel prediktornya merupakan faktor-faktor yang mempengaruhi tingkat pengangguran terbuka empat provinsi di Pulau Sumatera pada tahun 2020 dan dibatasi hanya lima variabel. Variabel tersebut terdiri dari jumlah penduduk, laju pertumbuhan penduduk, pendidikan, IPM dan PDRB.

1.4 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penelitian tugas akhir ini yaitu diperoleh faktor - faktor yang berpengaruh signifikan terhadap tingkat pengangguran terbuka empat provinsi di Pulau Sumatera pada tahun 2020.

1.5 Sistematika Penulisan

Penulisan penelitian ini terdiri dari lima bab. Pada Bab I Pendahuluan, dipaparkan hal-hal yang melatar belakangi penelitian ini, rumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan. Teori dan konsep yang akan digunakan untuk menyelesaikan masalah pada penelitian ini dijelaskan pada Bab II Landasan Teori. Selanjutnya, penelitian ini dilakukan dengan menggunakan data dan metode sebagaimana dijelaskan pada Bab III Metode Penelitian. Pada Bab IV Hasil dan Pembahasan dipaparkan deskripsi data dan analisis model yang diperoleh. Terakhir, kesimpulan dari penelitian ini dijelaskan pada Bab V dan ditutup dengan saran.