

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit *Coronavirus Disease* (Covid-19) adalah penyakit menular yang disebabkan oleh *Coronavirus* jenis baru yang ditemukan pada tahun 2019. Virus ini dapat menginfeksi manusia yang menyebabkan penyakit dengan gejala ringan sampai berat[36]. Sebelumnya ada enam jenis *Coronavirus* yang dapat menginfeksi manusia yaitu *Severe Acute Respiratory Illness Coronavirus* (SARS-CoV), *Alphacoronavirus 229E*, *Alphacoronavirus NL63*, *Betacoronavirus OC43*, *Betacoronavirus HKU1*, dan *Middle East Respiratory Syndrome Coronavirus* (MERS-CoV). Ada dua jenis *Coronavirus* yang dapat menyebabkan gejala berat yaitu *Middle East Respiratory Syndrome* (MERS) dan *Severe Acute Respiratory Syndrome* (SARS). *International Committee on Taxonomy of Viruses* menetapkan SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*) sebagai nama *Coronavirus* yang menyebabkan Covid-19, hal ini dikarenakan *Coronavirus* ini memiliki struktur genom yang mirip dengan SARS [29].

Covid-19 dapat menyebabkan infeksi sistem pernafasan, pneumonia dan lebih lanjut menyebabkan kematian[17]. Jumlah kasus Covid-19 kian hari semakin bertambah seiring berjalannya waktu. Menurut KPC PEN (Komite Penanganan Covid-19 dan Pemulihan Ekonomi Nasional) per April 2021, sudah

ada 1.566.995 jiwa yang terinfeksi di Indonesia. Dari total tersebut terdapat 1.414.507 kasus yang dinyatakan sembuh dan total kematian mencapai 42.530 jiwa [19]. Melihat masih tingginya angka kematian pasien Covid-19 di Indonesia, hal ini perlu dilakukan kajian lebih lanjut untuk mengetahui faktor apa saja yang dapat menyebabkan kematian pada pasien Covid-19. Salah satu metode yang dapat digunakan adalah metode klasifikasi.

Metode klasifikasi merupakan suatu metode yang dilakukan untuk mengelompokkan data sesuai dengan ciri-ciri atau karakteristik data tertentu [34]. Variabel respon dalam penelitian ini adalah status pulang pasien, dimana ada dua kemungkinan yaitu pasien dinyatakan sembuh atau meninggal. Variabel ini mengikuti distribusi Bernoulli, sehingga metode klasifikasi yang dapat digunakan diantaranya adalah metode *Naïve Bayes Classifier* (NBC) dan Regresi Logistik Biner (RLB).

Metode NBC merupakan metode klasifikasi yang memanfaatkan perhitungan peluang dan statistik dengan teorema Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [35]. Adapun regresi logistik adalah metode klasifikasi dengan analisis data statistika yang bertujuan untuk mengetahui hubungan kausal antara beberapa variabel, dimana variabel responnya bersifat kategori dan variabel prediktornya dapat bersifat kategori maupun kontinu. Regresi logistik dengan variabel responnya hanya terdapat dua kategori atau biner disebut dengan regresi logistik biner [3].

Beberapa penelitian yang dilakukan oleh Xhemali [35] dan Utomo [32] menyatakan bahwa metode klasifikasi NBC adalah metode klasifikasi yang lebih

baik diantara beberapa metode klasifikasi lainnya. Kemudian penelitian tentang metode ini juga pernah dilakukan oleh Derisma [9] yang menghasilkan bahwa metode klasifikasi NBC merupakan metode klasifikasi lebih baik dibandingkan metode *Random Forest* dan *Neural Network*. Sementara itu, penelitian yang dilakukan oleh Purwa [25] menyatakan bahwa metode klasifikasi regresi logistik memiliki hasil yang lebih baik dibandingkan dari metode lainnya. Penelitian oleh Salim [26], Seka [27] dan Fajrila [10] membandingkan metode klasifikasi NBC dan regresi logistik. Dalam penelitiannya dihasilkan bahwa metode NBC memiliki tingkat akurasi yang lebih tinggi dibandingkan regresi logistik. Namun, penelitian yang dilakukan oleh Prabowo [24] menyatakan bahwa metode regresi logistik dan NBC menghasilkan ketepatan klasifikasi atau tingkat akurasi yang sama. Perbedaan hasil analisis ini disebabkan karena data yang digunakan oleh peneliti berbeda satu sama lain.

Pendugaan parameter regresi dapat dilakukan dengan beberapa cara yaitu dengan metode klasik dan metode Bayes. Penelitian yang dilakukan oleh Azhar [1] dan Khairiyah [18] menyatakan bahwa pendugaan parameter dengan pendekatan Bayes adalah pendugaan parameter yang lebih baik dibandingkan metode klasik seperti metode Kuadrat Terkecil dan Maksimum *Likelihood*. Atas dasar itulah peneliti tertarik untuk membandingkan metode Regresi Logistik biner dengan pendugaan parameter menggunakan metode Bayes yang selanjutnya disebut Regresi Logistik Biner Bayesian (RLBB) dengan *Naïve Bayes Classifier* (NBC).

RLBB adalah regresi logistik biner dengan penduga parameter meng-

gunakan metode Bayes. Pada metode Bayes pendugaan parameter dilakukan dengan memanfaatkan informasi data atau fungsi *likelihood* dan distribusi prior untuk mendapatkan distribusi posterior. Nilai dugaan parameter diperoleh dari *mean* dari distribusi posterior atau disebut dengan *mean* posterior.

Berdasarkan uraian di atas yang menjadi fokus pada penelitian ini adalah perbandingan metode klasifikasi antara RLBB dan NBC. Kedua metode klasifikasi ini dipilih karena kedua metode tersebut merupakan metode klasifikasi yang lebih baik diantara beberapa metode klasifikasi lainnya. Kedua metode klasifikasi tersebut diaplikasikan pada data kasus terkait klasifikasi resiko kematian pasien Covid-19.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, adapun rumusan masalah dalam penelitian ini adalah:

1. Apa saja faktor-faktor yang mempengaruhi resiko kematian pasien positif Covid-19 dengan metode RLBB dan metode NBC ?
2. Bagaimanakah ketepatan klasifikasi yang didapatkan dengan menggunakan metode RLBB dan metode NBC ?

1.3 Tujuan Penelitian

Dari perumusan masalah di atas, tujuan dari penelitian ini adalah:

1. Mengetahui faktor-faktor yang mempengaruhi kematian pasien positif Covid-

19 dengan metode RLBB dan metode NBC.

2. Mengetahui ketepatan klasifikasi yang didapatkan dengan menggunakan metode RLBB dan metode NBC.

1.4 Batasan Masalah

Penelitian ini hanya mengklasifikasikan resiko kematian pada pasien Covid-19 di Sumatera Barat dengan metode RLBB dan NBC. Penelitian ini menggunakan tiga variabel prediktor yang diduga mempengaruhi resiko kematian pasien Covid-19 yaitu Jenis Kelamin, Umur dan Jumlah Komorbid. Data yang digunakan adalah data dari bulan Mei - Desember 2020 yang diambil dari Rumah Sakit Universitas Andalas dan Rumah Sakit M. Djamil, Sumatera Barat.

1.5 Sistematika Penulisan

Penelitian ini disusun dengan sistematika penulisan yang terdiri dari 5 (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 kematian pasien Covid-19, hasil klasifikasi faktor-faktor yang mempengaruhi kematian pasien Covid-19 dengan menggunakan metode RLBB dan metode NBC,

hasil keakuratan metode klasifikasi tersebut dan penentuan metode terbaik dalam klasifikasi faktor-faktor yang mempengaruhi kematian pasien Covid-19. Bab V Penutup, memberikan kesimpulan berdasarkan hasil yang diperoleh pada pembahasan dan juga disampaikan saran yang menjadi pedoman untuk peneliti selanjutnya.

