

BAB I

PENDAHULUAN

1.1. Latar Belakang

Covid-19 ialah penyakit respirasi akut yang diakibatkan oleh virus corona tipe baru. Penyakit ini mula-mula merebak di Wuhan China, kemudian menyebar hampir ke seluruh dunia sehingga menimbulkan adanya pandemi global. Indikasi penyakit Covid-19 antara lain demam, batuk serta sesak napas. Penyakit ini melanda seluruh kalangan, baik dewasa, lanjut usia, maupun kanak-kanak (Sari 2020). Covid-19 dapat menimbulkan gangguan ringan pada sistem respirasi, peradangan paru-paru yang berat, serta kerusakan paru-paru secara permanen, hingga kematian (Nasution 2020). Berdasarkan sebaran informasi dari Gugus Tugas Percepatan Penanganan Covid-19 di Indonesia pada tanggal 27 November 2020 tercatat total kasus mencapai 60,534,526 di seluruh dunia dan 522,581 kasus di Indonesia. Kenaikan kasus yang tidak bisa diperkirakan perkembangannya menyebabkan buruknya berbagai aspek kehidupan masyarakat Indonesia. Melalui penelitian ini diharapkan bisa menganalisis covid-19 berdasarkan data yang ada menggunakan algoritma regresi linier dan *random forest*.

Penelitian yang dilakukan Wiguna *et.al* (2020) mengenai Kebijakan Berbasis Data: Analisis dan Prediksi Penyebaran COVID-19 di Jakarta menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) memperoleh *Mean Absolute Percentage Error* (MAPE) sebesar 20,97%. Kemudian Albana dan Azhari (2020) mengenai prediksi penyebaran covid-19 kota Surabaya menggunakan metode simulasi monte carlo menghasilkan tingkat galat berupa *under-prediction* sebesar 11% dan *over-prediction* sebesar 23%. Sedangkan penelitian yang dilakukan Hardiyanti and Shofiyah (2020) mengenai prediksi kasus covid-19 di Indonesia menggunakan metode *adaptive neuro fuzzy inference system* (ANFIS) menghasilkan nilai presentase error sebesar 0.437571%. Selain itu Teguh, Sahay, dan Adji (2020) menjelaskan pemodelan penyebaran infeksi covid-19 di Kalimantan menggunakan metode model SIR kompartemen, menghasilkan pola pergerakan sebaran berbasis data aktual dengan kombinasi model SIR menjadi sangat penting untuk mendapatkan prediksi yang akurat. Arianto and Noviyanti (2020) mengemukakan mengenai prediksi kasus

covid-19 menggunakan metode *backpropagation* dan *fuzzy tsukamoto*, dengan metode tersebut diperoleh nilai koefisien korelasi atau $R = 0,84278$ dan simulasi prediksi memperoleh nilai MSE untuk data normalisasi adalah 1,632337. Kemudian hasil penelitian Pratikto (2020) mengenai prediksi akhir pandemi covid-19 di Indonesia dengan simulasi berbasis model pertumbuhan parametrik menggunakan metode *nonlinear least squares*, *k-fold cross validation* dan *bootstrap* memperoleh hasil simulasi tingkat keyakinan sebesar 95%.

Random forest ialah metode yang bisa meningkatkan hasil akurasi dan dapat menaikkan simpul anak untuk setiap node dengan secara acak. Metode ini berfungsi untuk membuat pohon keputusan yang mengambil atribut dan data secara acak berdasarkan ketentuan (Siburian dan Mulyana 2018). Penelitian yang dilakukan Siburian dan Mulyana (2018) mengenai prediksi harga ponsel menggunakan metode *random forest* menghasilkan nilai akurasi yang baik yaitu 81 %. Sedangkan, Regresi linear merupakan alat perlengkapan statistik yang digunakan untuk mengetahui pengaruh antara satu atau beberapa variabel terhadap satu buah variabel. Variabel yang mempengaruhi disebut variabel bebas, variabel independen atau variabel penjelas. Sedangkan variabel yang dipengaruhi disebut dengan variabel terikat atau variabel dependen. Regresi linear dapat digunakan pada skala interval maupun ratio (Kusumawati, Marisa, dan Wijaya 2017). Kemudian hasil penelitian Muhajir, Sutoyo dan Darmawan (2019) mengenai *Forecasting* model penyakit demam berdarah *dengue* di provinsi DKI Jakarta menggunakan algoritma regresi linier untuk mengetahui kecenderungan nilai variabel prediktor terhadap peningkatan kasus menghasilkan variabel yang berpengaruh terhadap kenaikan kasus adalah kelembapan, curah hujan dan temperatur.

Penelitian yang dilakukan Siregar *et al.* (2020) mengenai perbandingan algoritma klasifikasi untuk prediksi cuaca menggunakan algoritma *naive bayes*, *decision tree* dan *random forest* memperoleh nilai akurasi *naive bayes* sebesar 72,22%, *decision tree* 54,01% dan algoritma *random forest* sebesar 82,38%. Keberhasilan algoritma dapat dilihat dari seberapa besar nilai akurasinya, karena semakin tinggi nilai akurasi maka semakin baik algoritma tersebut dan semakin kecil memperoleh error. Dari ketiga algoritma tersebut *random forest* merupakan algoritma yang mempunyai nilai akurasi tertinggi.

Dari permasalahan di atas, maka dilakukan penelitian mengenai prediksi covid-19 di Indonesia menggunakan algoritma regresi linier dan *random forest* untuk mengetahui hasil prediksi dengan akurasi yang baik.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah di atas, adapun masalah yang akan dibahas adalah:

1. Bagaimana cara menganalisis kasus covid-19 di Indonesia menggunakan algoritma *random forest* dan regresi linier.
2. Bagaimana tingkat akurasi kasus covid-19 di Indonesia menggunakan algoritma *random forest* dan regresi linier.

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, adapun tujuan penelitian adalah sebagai berikut:

1. Mengetahui cara menganalisis kasus covid-19 di Indonesia menggunakan algoritma *random forest* dan regresi linier untuk mengetahui analisis terbaru.
2. Mengetahui tingkat akurasi algoritma *random forest* dan regresi linier dalam melakukan analisis kasus covid-19 di Indonesia.

1.4 Manfaat

Adapun manfaat penelitian adalah sebagai berikut:

1. Masyarakat dapat mengantisipasi kondisi yang akan terjadi ketika sudah mengetahui hasil analisis.
2. Membantu tenaga medis dalam mengantisipasi kebutuhan dan layanan kesehatan yang harus dipersiapkan ketika sudah mengetahui hasil analisis.
3. Membantu pemerintah dalam mempersiapkan strategi yang tepat untuk menangani permasalahan agar tidak terjadi lonjakan kenaikan kasus covid-19 di Indonesia.
4. Mengimplementasikan data mining untuk melakukan analisis dalam bidang kesehatan.