

BAB I PENDAHULUAN

1.1 Latar Belakang

Virus COVID-19 menjadi topik yang sangat hangat pada dua tahun terakhir ini di dunia. *World Health Organization* (WHO) menyatakan bahwa, COVID-19 merupakan suatu kelompok virus yang dapat menyebabkan infeksi pada saluran nafas pada manusia dan ditetapkan menjadi pandemi di seluruh dunia (Freni et al., 2020). COVID-19 dapat menular melalui droplet yang dikeluarkan oleh mulut dan hidung pada saat orang yang terinfeksi virus ini sedang batuk, bersin, berbicara lalu menempel pada benda lain seperti gagang pintu, meja, pakaian dan lain sebagainya dalam jarak 1 meter (Wiguna et al., 2020). Virus yang pertama kali di temukan di Wuhan, China ini telah menyebar hampir ke seluruh dunia termasuk Indonesia dengan sangat cepat (Regresi et al., 2022). Dengan sifatnya yang mudah menular, membuat Indonesia yang memiliki tingkat kepadatan penduduk tertinggi ke 4 di dunia menurut *United State Census Bureau*(1 Juli 2021), menjadi wilayah yang berpotensi tinggi terancam virus COVID-19 ini (United Census Bereau, 2022). Untuk saat ini menurut data yang dikutip dari laman www.covid.go.id Indonesia(26 April 2022), angka kasus Covid-19 di Indonesia mencapai 6 juta kasus.

Terdapat beberapa faktor yang menyebabkan tingginya angka kasus COVID-19, salah satunya adalah jumlah orang yang terkonfirmasi sebagai pasien positif aktif (*active cases*) COVID-19 (Elen Riswana Safila Putri et al., 2021). Menurut pemaparan Prof. Wiku Adisasmito, juru bicara satgas COVID-19 Indonesia saat konfrensi pers yang diadakan pada 13 Agustus 2021, *active cases* merupakan sebutan untuk seseorang yang dinyatakan positif COVID-19 dan masih dalam masa perawatan. Untuk menekan jumlah kasus COVID-19, perlu adanya deteksi dini dan diagnosa terhadap data yang diperoleh (MINARNO et al., 2021). Hal ini dapat dilakukan dengan membangun model *machine learning* regresi yang tepat untuk memprediksi angka *new active cases* COVID-19 (Rath et al., 2020).

Penggunaan *machine learning* yang berbasis model regresi untuk menyelesaikan permasalahan prediksi terhadap data yang bersifat *time series* sudah banyak dilakukan oleh beberapa penelitian sebelumnya. Seperti, penelitian yang dilakukan untuk memprediksi hasil produksi padi yang diimplementasikan ke dalam sebuah *website* (Ervan Triyanto, Heri Sismoro, 2019). Hasil pengujian yang dilakukan didapatkan bahwa MAD (*mean absolute deviation*) sangat kecil yaitu 0,040 dengan rata - rata kesalahan pada setiap data sebesar 0,101. Penelitian lain (Amrin, 2016) menggunakan *multiple linear regression* untuk memprediksi tingkat inflasi umum berdasar 7 variabel independen. Hasil performa model yang didapat dari penelitian ini yaitu MAD = 0.0380, MSE = 0.0023, dan RMSE = 0.0481, sehingga dapat disimpulkan bahwa akurasi dari model yang dibangun cukup akurat. Penelitian lainnya (Argawu et al., 2021) yang memprediksi *new cases* COVID-19 menggunakan analisis *person's correlation* dan *multiple linear regression* dilakukan di Ethiopia. Dari analisis *person's correlation* didapatkan beberapa 6 *feature* yang dapat mempengaruhi *new cases*, sehingga menghasilkan nilai R2 sebesar 96%.

Penelitian oleh Susanti, (2014) yang menggunakan algoritma *neural network* pada model regresi, dilakukan untuk memprediksi harga ayam menggunakan ANN berbasis *backpropagation*. ANN berbasis *backpropagation* dinilai optimal dan cukup akurat dengan menggunakan parameter toleransi *error* 0.001, learning rate 0.05 serta maksimum *epoch* 5000, serta didapatkan hasil nilai MSE sebesar 0.0113. Penelitian selanjutnya oleh NYONI et al., (2020) yang dilakukan di Brazil menyatakan bahwa model ANN memadai, stabil dan cocok digunakan untuk melakukan peramalan *new cases* harian COVID-19 di Kuwait, Brazil. Pada penelitian Hamdani & Setyanto, (2020), dilakukan perbandingan antara algoritma *neural network* dan *logistic regression* dalam melakukan prediksi terhadap jumlah mahasiswa yang dapat lulus tepat waktu (8 semester). Dalam penelitian ini, berdasarkan pengujian *K-fold Validation* diperoleh bahwa, *neural network* dinilai lebih unggul dengan selisih nilai akurasi sebesar 3% dari *logistic regression*.

Berdasarkan beberapa penelitian yang sudah disampaikan, penelitian ini membangun dan membandingkan model *machine learning* untuk mengetahui model mana yang lebih baik dalam memprediksi angka *new active cases*, yang selanjutnya akan disebut dengan model prediksi. Model ini dibangun menggunakan algoritma *Multiple Linear Regression* dan *Neural Network*. Kedua algoritma tersebut dipilih, karena masuk ke dalam algoritma yang cocok digunakan untuk solusi model prediksi pada data numerik yang bersifat *time series* dan memiliki lebih dari satu variabel independen (Masruroh, 2020). Penggunaan algoritma *Neural Network* pada dataset COVID-19 pernah dilakukan oleh beberapa penelitian. Seperti, penelitian yang dilakukan oleh Hendrawan et al., (2020), dimana algoritma *Neural Network* digunakan untuk melakukan prediksi jumlah kasus positif COVID-19 di Kota Semarang. Penelitian lainnya mengenai prediksi kasus COVID-19 menggunakan algoritma *Neural Network* adalah penelitian yang dilakukan oleh Niazkar & Niazkar, (2020). Pada penelitian tersebut, algoritma *Neural Network* dikatakan dapat digunakan untuk melakukan prediksi pada data *time series* dengan jumlah data yang cukup besar.

Dalam membangun sebuah model prediksi yang menerapkan model analisis regresi, diperlukan *dataset* yang merupakan sekumpulan data yang diolah agar dapat menghasilkan pengetahuan yang berguna berdasarkan tujuan dari dibangunnya model prediksi (Wahyudin & Purwanto, 2021). Dalam penelitian ini, akan dilakukan pula pemilihan *feature* yang memiliki korelasi positif terhadap *label* dan *tuning* terhadap *hyperparameter* dari masing - masing model. Kemudian, dilakukan analisis untuk mengetahui model dengan algoritma mana yang memiliki hasil yang lebih akurat. Pembangunan model prediksi dan analisa perbandingan yang dilakukan pada penelitian ini diharapkan dapat memberikan informasi yang berguna dalam mengklarifikasi rencana tindakan potensial serta mempersiapkannya apabila nantinya terjadi kenaikan ataupun penurunan *new active cases*.

1.2 Perumusan Masalah

Berdasarkan dari uraian penjelasan latar belakang permasalahan yang ada, maka perumusan masalah pada penelitian ini adalah:

1. Bagaimana mendapatkan *feature* yang mempengaruhi nilai *new active cases* pada dataset Persebaran COVID-19 di Indonesia?
2. Bagaimana membandingkan akurasi dari model prediksi *new active cases* pada dataset Persebaran COVID-19 di Indonesia, yang dibangun menggunakan *Multiple Linear Regression* dan *Neural Network*?

1.3 Tujuan Penelitian

Adapun tujuan penelitian yang dilakukan adalah untuk mendapatkan *feature – feature* yang mempengaruhi *label new active cases*, serta menganalisa dan membandingkan nilai akurasi yang diperoleh dari model yang dibangun dengan menggunakan algoritma *Multiple Linear Regresion* dan *Neural Network* pada prediksi angka *new active cases* COVID-19 di Indonesia, sehingga dapat diketahui model regresi dengan algoritma mana yang lebih baik untuk kasus yang diangkat pada penelitian ini.

1.4 Pembatasan Masalah

Adapun batasan masalah pada penelitian ini :

1. *Feature* yang digunakan untuk memprediksi *label* yang dimaksud dalam penelitian ini ditentukan berdasarkan hasil dari matriks korelasi (*heatmap*).
2. *Tools* yang digunakan untuk membangun model prediksi pada penelitian ini adalah *Rapidminer*.
3. *Dataset* yang digunakan pada penelitian ini adalah *dataset* Persebaran COVID 19 di Indonesia dari Klinik Mari Sehat.

1.5 Sistematika Penulisan

Adapun sistematika dari penulisan skripsi ini terdiri dari :

1. Bab I Pendahuluan, berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penulisan.

2. Bab II Tinjauan Pustaka Dan Landasan Teori, berisi tentang penelitian terdahulu dan teori-teori dasar mengenai *machine learning*, model regresi, *Cross Validation*, *Multiple Linear Regression*, *Neural Network*, Uji T dan *RapidMiner*.
3. Bab III Metode Penelitian, berisi tentang diagram alir dan rangkaian alur tahapan pelaksanaan penelitian, dari mulai tahap pengambilan dataset, analisis dataset, pembangunan model, evaluasi, analisis hasil serta pengambilan kesimpulan.
4. Bab IV Analisis Hasil Dan Pembahasan, berisi tentang hasil penelitian yang sudah dilakukan, serta pada bab ini juga dilakukan analisis hasil pengujian.
5. Bab V Penutup, berisi tentang kesimpulan terkait penelitian yang sudah dilakukan dan saran untuk melakukan perbaikan atau pengembangan pada penelitian lanjutan.