

BAB I PENDAHULUAN

1.1 Latar Belakang

Pada tahun 2019, wabah Coronavirus atau COVID-19 terdeteksi pertamakali di Wuhan, China. Kondisi seperti ini sangat mengkhawatirkan mengingat COVID-19 sangat mudah menular dan menyebabkan kematian. *Coronavirus* atau dikenal juga dengan *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) menurut *World Health Organization* (WHO) menjangkiti manusia yang menyebabkan infeksi pernapasan mulai dari flu biasa hingga penyakit yang lebih parah. Kasus COVID-19 pertama mulai masuk ke Indonesia pada 2 Maret 2020 yang telah diumumkan secara langsung oleh Presiden Joko Widodo [1]. COVID-19 ini menyebar dengan hebat ke seluruh dunia, menyebabkan banyak kematian dan kesulitan keuangan. COVID-19 telah mengorbankan setiap aspek kehidupan manusia dan menjadi beban berat bagi rumah sakit dan tenaga medis. Pada tahun 2020, penularan virus COVID-19 sulit dikendalikan yang mengakibatkan jumlah kasus COVID-19 meningkat sehingga pemerintah memberlakukan PSBB dan PPKM. Di Indonesia telah melakukan upaya untuk menekan angka penularan COVID-19 di masyarakat dengan kampanye 3M diantaranya (1) Mencuci tangan, (2) Memakai masker, dan (3) Menjaga jarak sebagai satu paket protokol kesehatan yang sangat diperlukan [2] hingga pada 2021 COVID-19 mulai menurun.

COVID-19 menurut *World Health Organization* (WHO) secara umum ditandai dengan gejala demam, batuk, kelelahan, dan kehilangan penciuman dan rasa. Deteksi dan observasi secara tepat pada pasien dinilai krusial untuk mengatasi pandemi yang mematikan ini. Mendiagnosis seseorang dari COVID-19 bermacam-macam seperti Swab PCR, Swab Antigen, Rapid test antibodi. Pada era digital ini, *machine learning* berperan yang menjanjikan dalam kondisi ini. *Machine learning* sendiri menurut Issam El Naqa et al. adalah bidang algoritma komputer yang terus berkembang dirancang untuk meniru kecerdasan manusia yang dapat digunakan dalam berbagai bidang mulai dari keuangan, hiburan, biologi komputasi hingga

medis [3]. Ada beberapa teknik yang dimiliki oleh *machine learning*, namun secara umum *machine learning* memiliki dua teknik dasar yaitu *supervised* dan *unsupervised*. *Unsupervised learning* menggunakan algoritma untuk menganalisis dan mengelompokkan dataset yang tidak berlabel [4], sedangkan *supervised learning* merupakan pendekatan *machine learning* yang ditentukan oleh dataset berlabel yang biasa digunakan ke dalam klasifikasi atau prediksi hasil secara akurat. Beberapa algoritma yang termasuk kategori *supervised learning* antara lain *random forest*, *k-nearest neighbours*, dan algoritma lainnya.

Dalam memprediksi diagnosis COVID-19 berdasarkan gejala akan digunakan teknik klasifikasi yang menghasilkan model prediksi yang akurat dengan menggunakan data training gejala COVID-19 untuk terlebih dahulu mempelajari pola-pola yang terbentuk dengan menggunakan algoritma yang akan mempelajari pola yang ada. Algoritma menghasilkan fungsi pendekatan yang menghubungkan variabel input dan outputnya. Variabel input pada penelitian ini akan menerapkan seleksi fitur yang merupakan salah satu tahapan *preprocessing* klasifikasi dimana memilih fitur yang relevan yang mempengaruhi hasil klasifikasi [5]. Variabel output adalah akurasi diagnosis COVID-19 melalui gejalanya.

K-Nearest Neighbors merupakan *supervised learning* yang digunakan sebagai metode klasifikasi terhadap kumpulan data. Pengklasifikasi k-NN menyimpan semua contoh pelatihan dalam memori [6], sehingga hasil yang baru diklasifikasikan dipengaruhi oleh besaran nilai "k" dimana berapa banyak jumlah data yang terdekat dengan suatu objek pada data testing [7]. *K-Nearest Neighbor* memiliki beberapa keunggulan diantaranya: pelatihan yang sangat cepat, sederhana dan mudah dipelajari, dan efektif jika data pelatihan besar [8]. Metode klasifikasi lain yang digunakan dalam penelitian ini adalah metode *Random Forest*. Metode *Random Forest* merupakan salah satu metode klasifikasi dimana terdapat proses agregasi *decision tree* yang mampu menghasilkan kinerja yang baik dan kuat untuk analisis data [9]. Keunggulan dari *Random Forest* antara lain dapat menghasilkan error yang lebih rendah, memberikan hasil yang bagus dalam klasifikasi, dan dapat mengatasi data training dalam jumlah yang sangat besar secara efisien.

Berdasarkan latar belakang diatas, penelitian ini mengusulkan riset tentang klasifikasi gejala COVID-19 berdasarkan kriteria gejala menggunakan *K-Nearest Neighbor* (K-NN) dan *Random Forest* (RF) dengan dataset yang diambil melalui situs kaggle.

1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian berdasarkan uraian latar belakang sebagai berikut:

1. Bagaimana perbandingan performa algoritma *K-Nearest Neighbor* (K-NN) dan *Random Forest* (RF) dalam mengklasifikasi gejala COVID-19?

1.3 Batasan Masalah

Agar pembahasan dalam penelitian ini dapat lebih jelas dan terarah maka penulis memberikan batasan terhadap permasalahan yang akan diteliti antara lain:

1. Data yang digunakan adalah data *Symptoms and COVID presence* melalui situs website Kaggle yang berjumlah 5434.
2. Menggunakan metode *K-Nearest Neighbor* (K-NN) dan *Random Forest* (RF) untuk melakukan pengklasifikasian.
3. Penelitian ini menggunakan pemrograman python dan software jupyterlab untuk melakukan proses analisis data.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengetahui perbandingan performa metode *K-Nearest Neighbor* (K-NN) dan *Random Forest* (RF) dalam mengklasifikasikan gejala COVID-19.

1.5 Manfaat Penelitian

Penelitian ini memiliki manfaat baik dari segi teoritis maupun praktis. Berikut adalah manfaat penelitian ini:

1. Manfaat Teoritis: Penelitian ini dapat memberikan kontribusi pada pengembangan pengetahuan dalam bidang klasifikasi gejala COVID-19. Hasil penelitian dapat menjadi referensi dan sumber informasi untuk

penelitian lanjutan yang berkaitan dengan metode klasifikasi penyakit.

2. Manfaat praktis: Hasil penelitian ini dapat digunakan sebagai dasar untuk mengembangkan metode diagnostik yang lebih akurat dalam mengklasifikasikan gejala COVID-19. Hal ini dapat membantu tenaga medis dalam melakukan diagnosis yang lebih tepat dan mempermudah proses penanganan pasien COVID-19.

1.6 Sistematika Penulisan

Secara garis besar penulisan laporan ini dibagi atas lima bab yang terdiri dari beberapa sub bab yaitu :

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab tinjauan pustaka ini meliputi telaah penelitian yang berisi tentang hasil peneliti yang sudah ada dan saling berkaitan dengan penelitian yang dilakukan dan landasan teori yang berisi tentang pembahasan mengenai pembahasan COVID-19, klasifikasi, seleksi fitur, *K-Nearest Neighbor* (K-NN) dan *Random Forest* (RF) yang diambil dari jurnal internet dan *e-book*.

BAB III METODE PENELITIAN

Dalam bab ini penulis mengemukakan tentang metode penelitian yang dilakukan oleh penulis dalam klasifikasi gejala COVID-19. Agar sistematis, bab metode penelitian meliputi tahapan pelaksanaan penelitian, pengolahan data (*pre-processing*), training dan testing data pada model, pembuatan model.

BAB IV HASIL PEMBAHASAN

Bab ini berisi tentang hasil penelitian yang telah dilakukan beserta pembahasannya terkait analisa hasil perhitungan yang digunakan dalam penelitian. Pembahasan ini meliputi pengukuran performa menggunakan *correlation matrix*, akurasi, ROC AUC, *precision*, *recall*, dan *F1 score*.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan dari hasil penelitian yang telah dilakukan dan berisi saran-saran yang dikemukakan penulis untuk pengembangan terhadap penelitian yang telah dilakukan.



