

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Deteksi dan diagnosis dini merupakan faktor penting untuk mengendalikan penyebaran COVID-19 saat ini, dimana pandemi besar COVID-19 saat ini telah memotivasi peneliti dalam menghasilkan solusi dengan teknologi digital berbantuan computer untuk diagnosis dan pencegahan COVID-19. Dalam mendiagnosis COVID-19 ada tiga teknik yang sering digunakan yaitu tes darah, X-ray, dan computed tomography (CT) scan. CT scan sampai saat ini merupakan teknologi yang sampai saat ini dianggap paling baik untuk mendeteksi COVID-19 (Hasan, AL-Jawad, Jalab, Shaiba, Ibrahim, & AL-Shamasneh, 2020) (Silva, et al., 2020). Menurut (Yang, He, Zhao, Zhang, Zhang, & Xie, 2020) banyak rumah sakit di Cina menggunakan CT untuk diagnosis COVID-19 dan terbukti efektif untuk diagnosis dini, namun CT tidak memiliki kemampuan untuk menentukan virus mana yang menyebabkan pneumonia virus: SARS-CoV-2 dan memastikan apakah seorang pasien terinfeksi COVID-19. Permasalahan saat ini adalah diagnosis COVID-19 di CT sangat bergantung pada ahlinya (Shuja, Alanazi, Alasmay, & Alashaikh, 2020), sehingga teknologi digital berbantuan computer seperti machine learning (ML) sangat diperlukan untuk efisiensi diagnosis dini COVID-19.

Machine learning (ML) didasarkan pada dimana sistem mampu mengenali pola dan melakukan tugas secara otomatis. Algoritma ML membentuk model matematika dari data sampel dan kemudian mengambil keputusan (Chahal & Gulia, 2019). Algoritma ML sangat penting bagi banyak masalah medis saat ini salah satunya adalah COVID-19, algoritma tersebut dapat membantu mendiagnosis penyakit, mendeteksi penyebabnya dan memprediksi hasil pengobatan, dll (Ala'raj, Majdalawieh, & Abbod, 2020). Dalam penelitian ini, akan diusulkan metode ML untuk mendeteksi citra x-ray positive dan negative COVID-19. Dataset Citra X-ray COVID-19 didalam penelitian ini diambil melalui data open akses dari <https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset>. Klasifikasi citra x-ray COVID-19 menggunakan ekstraksi fitur PCA dan metode

klasifikasi k-nearest neighbor (KNN) berbasis jarak euclidean dan manhattan. Tujuan penelitian ini yaitu untuk menganalisis algoritma jarak euclidean dan manhattan dalam mengklasifikasi citra x-ray COVID-19. Menurut (Al-Dosary, Al-Hamed, & Aboukarima, 2019) Algoritma K-NN adalah algoritma pembelajaran mesin yang dianggap algoritma pembelajaran yang malas, dengan biaya komputasi yang rendah dan implementasi yang sangat sederhana. Saat membuat prediksi, K-NN menyimpan seluruh set data pelatihan dan mengkueri untuk menemukan titik data k dalam set pelatihan yang paling mirip dengan titik data yang akan diklasifikasikan. Klasifikasi K-NN mempunyai beberapa kekurangan yaitu : kompleksitas perhitungan kesamaan data cukup besar performanya sangat dipengaruhi oleh sampel pelatihan tunggal dan K-NN tidak membentuk model klasifikasi karena K-NN adalah metode pembelajaran yang lambat (Aliwy & Ameer, 2017). Teknik kinerja K-NN adalah mencari jarak terpendek antar sampling untuk dievaluasi oleh nilai K terdekat dengan data pelatihan. Prinsip ini juga termasuk dalam bagian klasifikasi nonparametric. Dalam algoritma ini nilai K yang paling bagus dilihat dari datanya sehingga nilai K yang tinggi dapat berpengaruh pada pengurangan efek noise pada klasifikasi (Lubis, Sihombing, & Mawengkang, 2020). Fungsi jarak digunakan dalam klasifikasi K-NN adalah untuk memetakan sampel dengan kelas. jarak diterapkan antar sampel menggunakan metode euclidean atau manhattan atau minkowski (Raja & Nirmala, 2019). (ROSA, PRIMARTHA, & WIJAYA, 2019) Membandingkan Membandingkan algoritma KNN dengan pengukuran jarak untuk klasifikasi kanker serviks hasil percobaan menunjukkan bahwa yang paling cocok adalah jarak cosine karena memiliki nilai akurasi tertinggi 92,559% pada $k = 9$. (Ali, Neagu, & Trundle, 2019) Evaluasi kinerja pengklasifikasi tetangga terdekat k untuk kumpulan data yang heterogen hasil keseluruhan dari percobaan menunjukkan jarak euclidean bukanlah ukuran yang tepat yang bisa digunakan dengan KNN dalam mengklasifikasikan kumpulan data. (Baldini & Geneiatakis, 2019) menganalisis berbagai ukuran jarak untuk KNN dalam konteks deteksi malware OS Android, hasilnya menunjukkan bahwa pilihan pengukur jarak yang tepat dapat memberikan peningkatan yang signifikan pada akurasi klasifikasi. (Alfeilat, Hassanat, Lasassmeh, & Tarawneh, 2019) meneliti pengaruh pilihan pengukuran jarak pada K-Nearest dimana inti dari pengklasifikasi bergantung pada pengukuran jarak atau kesamaan antara contoh yang diuji dan contoh pelatihan. (Wahyono, Trisna, Sariwening, Fajar, & Wijayanto, 2020) Membandingkan euclidean, chebyshev, manhattan dan minkowski dengan dataset yang diambil melalui komentar pada youtube erninem , total dataset 488. Hasil uji coba yang dilakukan menunjukan jarak euclidean dan minkowski pada KNN memiliki akurasi lebih baik dibandingkan chebyshev dan

manhattan. Hasil klasifikasi pada KNN terbaik ketika K bernilai 3.

1.2 Rumusan Masalah

Berdasarkan latar belakang sebelumnya maka rumusan masalah yang dapat diambil adalah menganalisis penyakit COVID-19 pada citra x-ray menggunakan ekstraksi fitur PCA dan metode klasifikasi K-NN berbasis jarak euclidean dan manhattan untuk mendapatkan hasil negative dan positif penyakit COVID-19.

1.3 Tujuan dan Manfaat

1. Untuk mendapatkan perbandingan persentase keakurasian pada jarak euclidean dan manhattan dalam melakukan klasifikasi citra x-ray penyakit COVID-19 menggunakan ekstraksi fitur PCA dan metode klasifikasi K-NN.
2. Sebagai evaluasi perbandingan terhadap berbagai pendekatan lainnya.

1.4 Batasan Masalah

1. File citra x-ray yang digunakan di ambil dari : www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset. yang terdiri dari 2.482 citra x-ray COVID-19 dengan jumlah yang sama dari yang terinfeksi dan tidak terinfeksi COVID-19.
2. Menggunakan ekstraksi fitur PCA
3. Menggunakan metode klasifikasi K-NN

1.5 Keterbaruan

Pada penelitian-penelitian sebelumnya untuk deteksi penyakit COVI-19 menggunakan teknik tes darah, pemeriksaan citra x-ray dan CT scan .Dari hasil itu terdapat beberapa kesultan untuk deteksi penyakit COVID-19, terutama didalam mendiagnosis hasil citra x-ray COVID-19. Analisis yang

dilakukan selama ini dapat dikatakan subjektif, karena masih mengandalkan subjek yang menilai, dalam hal ini ahli medis. Dari permasalahan tersebut, maka diperlukan adanya suatu sistem analisis berbasis komputasi, seperti ML dimana sistem mampu mengetahui dari data, mengidentifikasi pola dan menciptakan keputusan serta melakukan tugas tertentu secara otomatis sehingga nilai yang di dapat lebih efisien dan objektif. Banyak penelitian deteksi COVID-19 sudah banyak dilakukan tetapi tidak ada yang membahas kinerja K-NN berbasis jarak dengan ekstraksi fitur PCA pada kasus mengklasifikasikan citra x-ray COVID-19.