

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Machine Learning (ML) atau pembelajaran mesin adalah cabang dari kecerdasan buatan yang memungkinkan sistem komputer untuk belajar dari data dan meningkatkan kinerjanya tanpa diprogram secara eksplisit. ML memanfaatkan algoritma dan model statistik untuk mengenali pola dalam data dan membuat prediksi atau keputusan berdasarkan pola tersebut [1]. *Machine Learning* dapat didefinisikan sebagai pemecahan masalah dalam memprediksi masa depan (*unobserved event*) atau memperoleh ilmu pengetahuan (*knowledge discovery*), serangkaian teknik yang dapat membantu dalam menangani dan memprediksi data yang sangat besar dengan cara mempresentasikan data-data tersebut dengan algoritma pembelajaran untuk meningkatkan performa atau membuat prediksi yang akurat [2]. Selain itu, dalam proses ML ini menggunakan metode klasifikasi.

Klasifikasi adalah metode dalam *machine learning* yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan karakteristik atau fitur yang dimiliki oleh data tersebut. Tujuan utama dari klasifikasi adalah untuk memprediksi label atau kategori dari data yang belum diketahui berdasarkan pola yang telah dipelajari dari data sebelumnya, dengan menerapkan algoritma klasifikasi yang tepat, berbagai sektor dapat meningkatkan efisiensi dan akurasi dalam pengambilan keputusan [3].

Ada banyak teknik klasifikasi yang dapat digunakan, diantaranya adalah *Support Vector Machine* (SVM) dan *Decision Tree* (C4.5). Konsep penelitian dengan algoritma SVM dan C4.5 telah banyak dilakukan oleh peneliti-peneliti sebelumnya, algoritma SVM dan C4.5 digunakan untuk klasifikasi data, pengenalan pola, deteksi, prediksi dan regresi [4].

Support Vector Machine (SVM) adalah salah satu algoritma *machine learning* yang digunakan untuk klasifikasi dan regresi. SVM bekerja dengan mencari *hyperplane* terbaik yang memisahkan data ke dalam kelas-kelas berbeda dengan margin atau jarak maksimal antar kelas tersebut. Algoritma ini juga dapat menangani data yang tidak dapat dipisahkan secara linier dengan menggunakan teknik *kernel trick* untuk memetakan data ke ruang berdimensi lebih tinggi sehingga memungkinkan pemisahan yang efektif [5]. Prinsip kerja SVM didasarkan pada konsep *Structural Risk Minimization* (SRM), yang bertujuan untuk menemukan *hyperplane* optimal yang memisahkan data dari dua kelas dengan margin terbesar, sehingga meningkatkan kemampuan generalisasi model terhadap data baru [6]. Data yang paling dekat dengan *hyperplane* disebut *support vectors*. Hanya data ini yang digunakan dalam menentukan posisi dan orientasi *hyperplane*, menjadikan SVM efisien dalam menangani data berdimensi tinggi [7].

Algoritma *Decision Tree* (C4.5) merupakan suatu algoritma yang dapat digunakan untuk melakukan klasifikasi atau pengelompokan pada dataset di mana algoritma ini membentuk pohon keputusan berdasarkan data yang tersedia untuk memetakan nilai-nilai atribut ke dalam label kelas tertentu [8]. C4.5 memilih atribut terbaik untuk dijadikan simpul (*node*) dalam pohon dengan menggunakan *Gain Ratio*, yaitu modifikasi dari *Information Gain*. *Gain Ratio* mengatasi kelemahan *Information Gain* yang cenderung memilih atribut dengan banyak nilai unik. Selain itu, C4.5 juga mampu menangani data numerik dan kategorikal, serta data yang memiliki *missing values*. Pohon yang dihasilkan oleh C4.5 dapat dilakukan proses *pruning* (pemangkasan) untuk mengurangi kompleksitas model dan menghindari *overfitting* [9]. SVM dan C4.5 adalah metode yang kuat dan efektif dalam klasifikasi, dan sering digunakan dalam berbagai penyakit, termasuk klasifikasi penyakit stroke, paru-paru, jantung dan ginjal kronis.

Dataset penyakit ginjal kronis didapat dari *Chronic Kidney Disease Dataset* [10], dataset ini mempunyai 1.659 jumlah data, 54 jumlah atribut serta 2 kelas yaitu 1 (terkena penyakit ginjal kronis) dan 0 (tidak terkena penyakit ginjal kronis). Ginjal merupakan salah satu organ terpenting pada manusia. Begitu banyak fungsi ginjal pada manusia seperti melakukan detoksifikasi darah dan menyaring limbah melalui urin. Ginjal juga berfungsi untuk menghilangkan kelebihan air pada tubuh atau menyimpan air saat dibutuhkan oleh tubuh. Terganggunya ginjal akan berpengaruh pada sistem dan fungsi ginjal. Ginjal yang tidak berfungsi dengan baik akan memberikan tanda-tanda pada tubuh, meskipun terkadang tidak terlalu dirasakan penderita [11].

Penyakit ginjal kronis merupakan penyakit dimana seseorang mengalami penurunan fungsi pada ginjal, ketika fungsi ginjal menurun mengartikan kondisi mengarah pada kerusakan ginjal. Penderita seringkali tidak merasakan gejala yang begitu terasa pada awal menderita penyakit ini. Sehingga penderita baru dapat terdiagnosa penyakit ginjal kronis ketika melakukan tes darah dengan indikasi-indikasi penyakit ginjal kronis yang lebih terlihat. Atau diagnosa ditegakkan sebagai hasil faktor resiko dari penyakit lain yang sedang diderita seperti tekanan darah tinggi dan diabetes [12]. Untuk mengidentifikasi penyakit ginjal kronis diperlukan pengklasifikasian menggunakan indikasi-indikasi yang menjadi acuan dalam diagnosa penyakit ginjal kronis. Proses klasifikasi ini penting karena dapat meningkatkan akurasi diagnosis dan mempermudah pengambilan keputusan medis. Metode klasifikasi juga digunakan dalam pengembangan sistem pendukung keputusan berbasis komputer dalam dunia medis [13].

Berdasarkan dari latar belakang diatas, maka peneliti akan melakukan pengujian terhadap algoritma *Support Vector Machine* (SVM) dan *Decision Tree* (C4.5) untuk

mengklasifikasi penyakit ginjal kronis serta mendapatkan metode mana yang memiliki nilai akurasi paling tinggi dengan berjudul “**Komparasi Algoritma Support Vector Machine (SVM) Dan Decision Tree (C4.5) Untuk Klasifikasi Penyakit Ginjal Kronis**”.

1.2 RUMUSAN MASALAH

Berdasarkan dari latar belakang yang telah diuraikan di atas, maka di dapat Rumusan masalah sebagai berikut :

1. Bagaimana penerapan *machine learning* dengan algoritma SVM dan C4.5 dalam mengklasifikasi penyakit ginjal kronis?
2. Bagaimana perbandingan akurasi klasifikasi antara algoritma SVM dan C4.5 dalam mendeteksi penyakit ginjal kronis?

1.3 BATASAN MASALAH

Untuk menghindari permasalahan yang meluas, maka penulis membatasi pembahasan permasalahan sebagai berikut:

1. Data yang digunakan bersumber dari *repository public* yaitu *kaggle.com* dengan nama *Chronic Kidney Disease Dataset*.
2. Metode klasifikasi serta algoritma yang digunakan yaitu *Support Vector Machine (SVM)* dan *Decision Tree (C4.5)*.
3. Kinerja algoritma SVM dan C4.5 diukur dengan menggunakan *Confusion Matrix* yang mencakup nilai *accuracy*, *precision*, *recall*, dan *F1-score*.
4. Aplikasi yang digunakan adalah *Google Colaboratory*.
5. Dataset penelitian menggunakan 1.659 data.
6. Bahasa pemrograman yang digunakan adalah *Python*.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut :

1. Untuk menerapkan *machine learning* dengan algoritma SVM dan C4.5 dalam mengklasifikasi penyakit ginjal kronis.
2. Mendapatkan nilai akurasi terbaik dari perbandingan antara algoritma SVM dan C4.5 dalam mengklasifikasi penyakit ginjal kronis.

1.4.2 Manfaat Penelitian

Adapun beberapa manfaat yang akan didapat dalam melakukan Penelitian ini, yaitu :

1. Dapat membantu mengetahui metode mana yang lebih akurat dan cocok untuk mendiagnosis pasien berdasarkan *dataset* penyakit ginjal kronis.
2. Dapat berkontribusi pada peningkatan pengetahuan medis terkait faktor risiko dan karakteristik pasien yang rentan terhadap penyakit ginjal kronis.
3. Memberikan kontribusi keilmuan pada penelitian bidang klasifikasi *machine learning* khususnya untuk klasifikasi penyakit ginjal kronis.
4. Dapat digunakan sebagai acuan untuk melakukan penelitian berikutnya.

1.5 SISTEMATIKA PENULISAN

Sistematika penulisan menggambarkan secara umum tentang apa yang akan dibahas dalam setiap bab dari laporan penelitian ilmiah ini. Adapun isi pokok bahasan masing-masing bab sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini berisi tentang latar belakang masalah, rumusan masalah, Batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab ini menjelaskan teori-teori yang mendasari pembahasan laporan termasuk penyakit ginjal kronis, konsep dasar *Machine Learning*, dan algoritma *Support Vector Machine* (SVM) dan *Decision Tree* (C4.5). Selain itu, disajikan pula tinjauan terhadap penelitian-penelitian terdahulu yang relevan.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan metode penelitian yang digunakan, mulai dari tahap pengumpulan data, *preprocessing*, pemilihan algoritma, evaluasi model, hingga validasi menggunakan teknik *cross validation*.

BAB IV : ANALISIS

Pada bab ini menjelaskan tentang Gambaran umum pemilihan dataset pada *Kaggle.com* kemudian proses penerapan algoritma SVM dan C4.5 pada dataset penyakit ginjal kronis, meliputi deskripsi data, tahapan *preprocessing*, implementasi model klasifikasi, evaluasi kinerja model, serta analisis perbandingan performa kedua algoritma.

BAB V : KESIMPULAN

Pada bab ini berisi kesimpulan yang diperoleh dari hasil penelitian serta saran untuk pengembangan penelitian di masa yang akan datang.