

BAB I

PENDAHULUAN

1.1 Latar Belakang

Liver atau hati merupakan salah satu organ yang sangat penting bagi manusia. Organ ini terletak di rongga perut bagian kanan, berada tepat di bawah diafragma. Umumnya hati sebagai organ memiliki fungsi sebagai penangkal dan penetral racun, mengatur siklus hormonal, mengatur komposisi darah yang mengandung zat lemak, gula, protein dan zat lainnya. liver juga bertanggung jawab untuk memproduksi zat empedu yang berfungsi untuk membantu pencernaan (Setiawati dkk., 2019).

Penyakit liver adalah salah satu penyakit yang sering terjadi di seluruh dunia. Penyakit ini dapat disebabkan oleh berbagai faktor seperti konsumsi alkohol, obesitas, dan infeksi virus hepatitis. Penyakit liver adalah penyakit yang mengganggu kinerja fungsi hati. Hati bertanggung jawab untuk fungsi kritis dalam tubuh sehingga ketika hati tidak menjalankan fungsinya dengan baik maka akan menyebabkan kerusakan serius pada tubuh. Hati adalah satu-satunya organ dalam tubuh yang dapat dengan mudah meregenerasi atau mengganti sel-sel yang rusak, tetapi jika hati rusak sehingga sel-sel itu tidak bisa diregenerasi atau hilang maka liver tidak mungkin bisa memnuhi kebutuhan tubuh (Patimah dkk., 2021). Penyakit liver sering kali dijuluki sebagai silent killer atau pembunuh diam-diam karena timbulnya penyakit hati sering kali tidak diawali dengan gejala atau yang sering dikenal dengan istilah asimptomatik. Penyakit liver dapat berdampak pada kesehatan secara keseluruhan dan dapat menyebabkan kerusakan organ dan bahkan kematian. Oleh karena itu, diagnosis dan pengobatan penyakit liver harus dilakukan dengan cepat dan tepat (Pusporani dkk., 2019).

Dalam bidang kesehatan, perkembangan teknologi informasi telah membantu dalam diagnosis dan pengobatan penyakit melalui penggunaan algoritma klasifikasi pada data medis. Salah satu algoritma klasifikasi yang sering digunakan adalah algoritma C5.0. Algoritma C5.0 merupakan salah satu algoritma pohon keputusan yang sering digunakan dalam klasifikasi data medis. Algoritma C5.0 merupakan penyempurnaan algoritma sebelumnya yang dibentuk oleh Ross Quinlan pada

tahun 1987, yaitu ID3 dan C4.5 yang mana Quinlan memberikan perbaikan pada algoritma C5.0 dibandingkan algoritma sebelumnya pada penggunaan memori yang lebih sedikit dan tree yang dihasilkan lebih ringkas. Dalam algoritma C5.0 langkah-langkah penyelesaiannya sangat mirip dengan algoritma C4.5 dimana kemiripan tersebut terdapat pada perhitungan *entropy* dan *gain*, perbedaannya pada algoritma C5.0 tidak hanya berakhir pada pencarian *gain* tapi juga dilakukan pencarian nilai *gain ratio* yang nantinya nilai *gain ratio* terbesar akan digunakan untuk menentukan cabang node selanjutnya. Algoritma ini bekerja dengan membuat pohon keputusan berdasarkan fitur yang ada pada data dan menghasilkan aturan-aturan yang dapat digunakan untuk mengklasifikasikan data (Wijaya dkk., 2018).

Namun, pada data medis kompleks, model-model rentan terhadap *overfitting*, sehingga diperlukan teknik khusus untuk membuat model-model tersebut lebih tahan terhadap *overfitting*. *Overfitting* terjadi ketika model terlalu kompleks dan mampu mengingat setiap detail pada data training, sehingga tidak dapat melakukan generalisasi pada data baru. Salah satu teknik yang dapat digunakan untuk mengatasi masalah ini adalah boosting dengan menggunakan algoritma Adaboost. Algoritma ini cocok untuk diterapkan pada dataset medis kompleks karena dapat meningkatkan akurasi model dan membuatnya lebih tahan terhadap *overfitting*. Salah satu keunggulan utama dari AdaBoost adalah kemampuannya untuk menangani masalah ketidakseimbangan kelas pada dataset (Tanti, 2023). Hal ini utamanya dicapai melalui penyesuaian bobot sampel yang dilakukan oleh AdaBoost. Dengan memberikan bobot yang lebih tinggi pada sampel-sampel yang salah diklasifikasikan, AdaBoost memberikan perhatian lebih pada kelas-kelas yang sulit diklasifikasikan, terutama kelas minoritas. Penelitian terdahulu mengenai klasifikasi suatu penyakit menggunakan algoritma C5.0 yang telah diboosting dengan Adaboost pernah dilakukan dengan judul “Diagnosa penyakit disk hernia dan spondylolisthesis menggunakan algoritma C5.0”. Penelitian tersebut berhasil mengklasifikasikan penyakit disk hernia dan spondylolisthesis dengan baik, akurasi yang didapatkan sebesar 79% dan akurasi bisa ditingkatkan menjadi 83% setelah di boosting menggunakan metode Adaboost (Saidata Aesy dkk., 2019).

Oleh karena itu pada penelitian ini peneliti akan menggunakan algoritma C5.0 yang diboosting dengan adaboost untuk melakukan klasifikasi pada data medis. Bila dibandingkan dengan penelitian pendahulu yang sama-sama memiliki data medis, pada penelitian ini dataset memiliki atribut yang lebih banyak dan keseimbangan dataset yang lebih ekstrem sehingga data yang digunakan lebih kompleks. Selain itu algoritma adaboost yang digunakan yaitu varian `adaboost.m1`, hal ini dikarenakan pada adaboost versi ini memang dirancang khusus untuk melakukan *binary classification*.

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan teknik klasifikasi pada data medis yang kompleks seperti penyakit liver dengan akurasi yang baik. Penelitian ini diharapkan dapat memberikan kontribusi pada bidang kesehatan dengan menyediakan sebuah model klasifikasi yang dapat membantu dokter dalam melakukan diagnosis liver dengan lebih cepat dan efektif. Selain itu, penelitian ini juga diharapkan dapat memberikan pemahaman yang lebih baik mengenai C5.0 dalam melakukan klasifikasi pada dataset medis.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas maka dapat dirumuskan permasalahan yang akan dibahas antara lain:

1. Bagaimana hasil ketepatan klasifikasi penyakit liver menggunakan algoritma C5.0 sebelum dilakukan boosting menggunakan algoritma adaboost ?
2. Bagaimana hasil ketepatan klasifikasi penyakit liver menggunakan algoritma C5.0 setelah dilakukan boosting menggunakan algoritma adaboost ?
3. Bagaimana perbandingan ketepatan hasil klasifikasi metode algoritma C5.0 sebelum dan sesudah dilakukan boosting menggunakan algoritma adaboost ?

1.3 Batasan Masalah

Adapun batasan masalah yang yang ditetapkan agar pembahasan yang terdapat pada penelitian ini tidak menyimpang adalah sebagai berikut:

1. Sumber data yang digunakan pada penelitian ini yaitu diagnosa penyakit liver yang disediakan oleh (UCI) machine learning university pada website Kaggle.

2. Hasil akhir dari pengujian berupa hasil evaluasi model menggunakan *confusion matrix* yang ditanamkan dalam sistem.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disampaikan, maka tujuan dari penelitian ini diantaranya:

1. Untuk memperoleh hasil ketepatan klasifikasi penyakit liver menggunakan algoritma C5.0 sebelum dilakukan boosting menggunakan algoritma adaboost.
2. Untuk memperoleh hasil ketepatan penyakit liver menggunakan algoritma C5.0 setelah dilakukan boosting menggunakan algoritma adaboost.
3. Untuk memperoleh perbandingan performa hasil klasifikasi metode algoritma C5.0 sebelum dan sesudah dilakukan boosting menggunakan algoritma adaboost.

1.5 Manfaat Penelitian

Manfaat yang didapatkan dari penelitian yang dilakukan diantaranya:

1. Bagi peneliti
Adapun manfaat dalam penelitian ini yaitu mengetahui perbandingan dari akurasi algoritma C5.0 sebelum dan sesudah dilakukan boosting menggunakan adaboost untuk mengklasifikasi penyakit liver
2. Bagi mahasiswa
Harapannya hasil dari penelitian ini dapat membantu mahasiswa menambah wawasan dan pengetahuan tentang metode algoritma C5.0 beserta teknik boosting menggunakan adaboost.
3. Bagi peneliti lain
Hasil dari penelitian ini diharapkan bisa menjadi referensi bagi peneliti yang ingin meneliti menggunakan algoritma C5.0 untuk mengklasifikasikan data, begitu pun juga bisa digunakan sebagai bahan referensi untuk mengembangkan penelitian lebih lanjut dibidang yang terkait.