

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Menurut WHO diperkirakan lebih dari 58 juta orang yang ada di dunia sudah terkena penyakit hepatitis C, dan sekitar 1,5 juta manusia terinfeksi penyakit hepatitis C setiap tahunnya. Salah satu yang penyebab meningkatnya jumlah kasus hepatitis C adalah keterlambatan diagnosis penyakit hepatitis C itu sendiri.

Hepatitis merupakan penyakit radang pada hati yang bisa disebabkan oleh berbagai kasus, termasuk infeksi virus. Terdapat 4 macam virus yang berhubungan dengan hepatitis yaitu, Hepatitis A, Hepatitis B, Hepatitis C, dan Hepatitis D.

Virus hepatitis C dianggap sebagai virus yang dapat menimbulkan dampak paling besar/mematikan diantara penyebab hepatitis lainnya. Beberapa atau banyak manusia tidak sadar bahwa telah terinfeksi virus hepatitis C hingga hati penderita telah rusak [1]. Salah satu jenis virus yaitu *RNA* adalah virus hepatitis C yang merupakan virus akan bertugas memberi sinyal dalam membuat *DNA*/ptotein baru. Ketika virus menginfeksi *RNA*, itu akan mengubah bentuk *DNA*. Virus hepatitis C (*HCV*) adalah virus *RNA* tunggal yang sudah resmi dianalisis pada bulan April tahun 1989 sebagai pemicu terpenting pada hepatitis non-A, dan non-B [2].

Oleh karena itu diperlukan *prognosis* (prediksi) dini untuk penyakit hepatitis C berdasarkan dari atribut yang terdapat pada dataset. Didalam melakukan diagnosis data dalam penelitian ini, metode yang di gunakan pada penelitian ini adalah Decision Tree Algoritma C4.5. Metode Decision Tree dengan menggunakan

Algoritma C4.5 dapat membuat prediksi berdasarkan dataset yang akan digunakan untuk menghitung probabilitas penyakit, dan juga dapat melihat seberapa efektif algoritma *Decision Tree* dalam mendeteksi penyakit Hepatitis C. Adapun alasan mengapa penulis memilih dan menggunakan Algoritma C4.5 (*Decision tree*) karena metode tersebut banyak digunakan oleh peneliti lainnya. Tidak hanya itu algoritma ini mampu menghasilkan akurasi yang sangat tinggi serta juga dapat membuat penelitian yang dilakukan lebih efisien dikarenakan dapat menghilangkan perhitungan yang tidak diperlukan.

Banyak Penelitian yang memprediksi penyakit hepatitis C dengan klasifikasi *data mining*, Diantara nya dilakukan oleh [3] hasil dari Algoritma C4.5 dibuat dengan metode yang lain yaitu Metode *Adaboost*. Nilai akurasi yang dihasilkan adalah sebesar 95,60%. Lalu terdapat juga penelitian [4] yang menggunakan Algoritma C4.5 dengan tingkat akurasi hasil 77,29%.

Berdasarkan Latar belakang yang telah dibuat, penulis tertarik untuk melakukan penelitian yang berjudul ***“Penerapan Data Mining Dengan Menggunakan Algoritma C4.5 Untuk Memprediksi Penyakit Hepatitis C”***.

1.2 RUMUSAN MASALAH

Berdasarkan dari latar belakang tersebut, maka didapatkanlah rumusan masalah sebagai berikut :

1. Bagaimana mendeteksi penyakit Hepatitis C menggunakan metode Algoritma C4.5?

2. Bagaimana hasil pohon keputusan yang didapat Algoritma C4.5 dalam memprediksi penyakit Hepatitis C?
3. Seberapa besar tingkat akurasi Algoritma C4.5 apabila digunakan untuk memprediksi penyakit Hepatitis C dengan 2 buah *tools*?

1.3 BATASAN MASALAH

Adapun batasan masalah yang terdapat dalam penelitian ini antara lain :

1. Data yang diperlukan bersumber dari *dataset* penyakit hepatitis C sebanyak 615 data. (<https://www.kaggle.com/datasets/fedesoriano/hepatitis-c-dataset>)
2. Variabel atau atribut yang terdapat pada dataset dan digunakan dalam penelitian ini adalah *Category*, *Age*, *Sex*, *ALB (Albumin Blood Test)*, *ALP (Alkaline phosphatase)*, *ALT (Alanine Transaminase)*, *AST (Aspartate Transaminase)*, *BIL (Bilirubin)*, *CHE (Acetylcholinesterase)*, *CHOL (Cholesterol)*, *CREA (Creatinine)*, *GGT (Gamma-Glutamyl Transferase)*, *PROT (Proteins)*.
3. *Tools* (alat bantu) yang digunakan untuk klasifikasi adalah *WEKA* dan *RapidMiner*

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Adapun tujuan dari penelitian yang akan dilakukan oleh penulis, yaitu :

1. Menerapkan Algoritma C4.5 untuk mengklasifikasi penyakit Hepatitis C.

2. Menghasilkan pohon keputusan yang tepat agar mudah dalam menentukan penyakit Hepatitis C
3. Mendapatkan akurasi yang tepat serta mengetahui tingkat akurasi dengan menggunakan Algoritma C4.5 dalam klasifikasi Hepatitis C.

1.4.2 Manfaat Penelitian

Adapun manfaat dalam penelitian ini, yaitu :

1. Dapat membantu memberikan kemudahan dan mampu memberikan solusi yang tepat dalam klasifikasi penyakit hepatitis C dalam dunia medis.
2. Mengetahui hasil pohon keputusan yang didapat dari 2 *tools* yaitu *WEKA* dan *RapidMiner* yang dapat dipahami dengan mudah.
3. Mengetahui tingkat akurasi yang lebih baik dari kedua *tools* yang berbeda yaitu menggunakan *weka* dan *rapid miner*.

1.5 SISTEMATIKA PENULISAN

Sistem penulisan ini secara umum menggambarkan apa yang penulis cakup dalam setiap bab dari laporan ini. Sistematika tugas akhir ini terdiri dari beberapa bab, yaitu:

BAB I : PENDAHULUAN

Bab ini akan menjelaskan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penulisan, serta sistematika atau struktur penulisan.

BAB II : LANDASAN TEORI

Bab ini memuat tentang landasan teori yang mendasari pembahasan laporan secara khusus dan berisikan topik permasalahan sesuai yang diteliti.

BAB III : METODOLOGI PENELITIAN

Bab ini membahas secara rinci kerangka kerja penelitian, metode atau pendekatan yang digunakan, dan *tools* (alat bantu) yang digunakan dalam menjawab masalah penelitian untuk mencapai suatu tujuan.

BAB IV : ANALISIS

Dalam bab ini akan menjelaskan dan memperlihatkan proses dari aktivitas perhitungan dan juga analisis yang dilakukan, kemudian pada tahap akhir akan menampilkan hasil yang telah diperoleh.

BAB V : KESIMPULAN

Pada bab terakhir berisi kesimpulan yang berkaitan dengan tujuan penelitian dan hasil yang telah diperoleh, serta berisi saran untuk penelitian selanjutnya.