BAB I

PENDAHULUAN

1.1 Latar Belakang

Hepatitis merupakan jenis penyakit menular dengan ciri utama yaitu peradangan atau infeksi pada organ hati [1]. Berdasarkan data dari WHO *(World Health Organization)* pada tahun 2019 diperkirakan terdapat 78.000 kematian di seluruh dunia akibat komplikasi yang timbul dari hepatitis A hingga E akut [2]. Tanggal 28 Juli diperingati hari hepatitis sedunia untuk meningkatkan kesadaran akan virus hepatitis, yang menyebabkan peradangan hati yang mengakibatkan penyakit parah dan kanker hati. Pada tanggal 05 April 2022 WHO menyampaikan bahwa telah menerima laporan mengenai 10 (sepuluh) kasus hepatitis akut yang belum diketahui sumber dan penyebabnya. WHO menetapkan hal tersebut sebagai Kejadian Luar Biasa (KLB), sejak dipublikasi secara resmi per tanggal 21 April 2022, tercatat 169 kasus yang dilaporkan di 12 negara yaitu Inggris (114), Spanyol (13), Israel (12), Amerika Serikat (9), Denmark (6), Irlandia (5), Belanda (4), Italia (4), Norwegia (2), Perancis (2), Romania (1) dan Belgia (1) [3]. Sehubungan dengan KLB tersebut WHO menyerukan kepada negara di seluruh dunia agar mencapai target yaitu, mengurangi kematian terkait hepatitis dan kanker sebesar 65% serta setidaknya 80% dari mereka yang memenuhi syarat menerima perawatan dan pengobatan yang tepat hal tersebut bertujuan untuk mencapai eliminasi pada tahun 2030 [4]. Hepatitis atau infeksi peradangan hati, menurut Dr. dr. Irsan Hasan Sp. PD-KGEH.FINASIM selaku Ketua Penelitian Hati Indonesia, adalah penyakit *silent killer* karena umumnya tanpa gejala, sehingga banyak orang yang tidak menyadari tengah menderita hepatitis [5]. Kendati memiliki ancaman kematian yang tinggi, guna menekan angka kenaikan penyakit hepatitis para pakar kesehatan harus melakukan diagnosis penyakit sedini mungkin. Dalam bidang kedokteran modern, pemeriksaan kesehatan terkait penyakit dapat dilakukan secara otomatis dengan bantuan teknologi informasi, yaitu dengan menggunakan pendekatan kerja data mining [6].

Data mining dideskripsikan sebagai teknik otomatis yang digunakan untuk mengeksplorasi secara menyeluruh pada set data yang kemudian akan diolah sebagai pembantu dalam pengambilan keputusan [7] [8]. Terdapat 5 (lima) metode pengelompokan dalam data mining yaitu klasifikasi, estimasi, prediksi, *clustering*, dan asosiasi [9]. Metode klasifikasi diambil sebagai metode yang akan digunakan dalam penelitian ini, karena metode tersebut banyak digunakan dalam dunia kesehatan salah satunya yaitu penyakit hepatitis [10]. Terdapat beberapa algoritme klasifikasi dalam data mining seperti *Support Vector Machine* (*SVM*) menurut penelitian [11] [12] [13] memiliki pendekatan *supervised learning* yang bekerja untuk memisahkan kelas dengan menentukan fungsi hyperplane atau pemisah yang ideal. K-*Nearest Neighbor (KNN)* menurut penelitian [14] [15] [8] merupakan salah satu algoritme klasifikasi berdasarkan kemiripan atau kedekatan terhadap data lainnya. C4.5 menurut penelitian [17] [18] [19] merupakan algoritme yang digunakan untuk melakukan klasifikasi dalam pembentukan pohon keputusan. *Naïve Bayes* menurut penelitian [20] dan [21] merupakan algoritme klasifikasi atau pengelompokan data dengan ciri utama mempunyai perkiraan yang kuat akan kekokahan dari masing-masing kondisi. *Random Forest* menurut penelitian [22] [23] [24] merupakan algoritme metode berbasis klasifikasi dan regresi yang melibatkan kumpulan pohon keputusan yang nantinya akan dijadikan *vote* untuk mendapatkan sebuah *output*.

Berdasarkan penelitian sebelumnya oleh Siti Khomsah yang berjudul Prediksi Harapan Hidup Penderita Hepatitis Kronik Menggunakan Metode-Metode Klasifikasi (2018) [25] didapat algoritme terbaik yaitu algoritme *Random Forest* [26] [27] [28]. Pada penelitian tersebut pengujian terhadap dataset menggunakan algoritme *Random Forest* hanya dilakukan sebanyak 1 (satu) kali dengan hasil nilai akurasi mencapai 79,35%. Algoritme *Random Forest* pada penelitian ini akan dilakukan pengujian terhadap *maximal depth* sebanyak 10 (sepuluh) kali dengan mencari kedalaman terbaik sehingga dapat membuat nilai akurasi semakin tinggi**.** Pada penelitian ini juga dilakukan perhitungan parameter waktu komputasi terhadap algoritme yang digunakan sehingga membuat proses validasi kinerja semakin baik. Algoritme *Random Forest* akan dibandingkan dengan algoritme

Support Vector Machine (*SVM*) [29] [30] [31]. Kelebihan algoritme *Random Forest* terletak pada hasil klasifikasi yang cenderung bagus dengan mengasilkan pohon keputusan lebih dari satu hal tersebut menjadikan hasil yang diperoleh lebih akurat, termasuk jenis algoritme klasifikasi yang dapat menangani *missing* data dengan baik [32]. Sedangkan kelebihan untuk algoritme *Support Vector Machine* (*SVM*) adalah cocok untuk ruang dimensi tinggi yaitu permasalahan dengan jenis parameter banyak, dengan waktu pemrosesan yang relatif singkat [33].

Berdasarkan latar belakang yang telah dijelaskan penelitian bertujuan mencari algoritme terbaik antara Algoritme *Random Forest* dan Algoritme *Support Vector Machine* (*SVM*). Manfaat yang diharapkan dalam penelitian ini ialah algoritme terbaik dapat digunakan sebagai dasar penelitian selanjutnya untuk proses pengklasifikasian penyakit hepatitis yang diharapkan dapat meminimalisir keterlambatan proses diagnosis sehingga dapat menekan kenaikan angka penyakit hepatitis serta dapat mengurangi resiko angka kematian pada pasien.

1.2 Rumusan Masalah

Diagnosis merupakan langkah awal penentu suatu jenis penyakit dengan memeriksa gejalanya. Berdasarkan latar belakang, perlu dilakukan klasifikasi terhadap penyakit hepatitis dikarenakan jenis penyakit ini memiliki angka kematian yang relatif tinggi dalam waktu yang relatif singkat. Karena ada banyak algoritme yang dapat menangani klasifikasi maka diperlukan perbandingan mencari algoritme terbaik guna mempercepat proses diagnosis yang diharapkan dapat meminimalisir angka kematian terhadap pasien penyakit hepatitis.

1.3 Pertanyaan Penelitian

Berdasarkan rumusan masalah tersebut, manakah diantara algoritme *Random Forest* dan *Support Vector Machine* (*SVM*) yang memiliki kemampuan terbaik dalam proses pengklasifikasian terhadap harapan hidup pasien hepatitis?

1.4 Batasan Masalah

Penelitian ini terbatas pada beberapa hal, diantaranya:

a. Data yang digunakan dalam penelitian ini berupa data public yang diambil dari UCI https://archive.ics.uci.edu/ml/datasets/Hepatitis dengan karakteristik data multivariasi bertipe kategoris, bilangan bulat dan nyata.
b. Data berjumlah 155 *record* dari 20 atribut, dengan atribut utama adalah *Class* yang menunjukan harapan hidup pasien *DIE or LIVE*.
c. Algoritme *Support Vector Machine* (*SVM*) dengan kernel *Radial Basis Function* (RBF).
d. Analisis kemampuan meliputi akurasi dan lama waktu komputasi.

1.5 Tujuan

Berdasarkan rumusan masalah, penelitian ini bertujuan untuk menganalisis dan memvalidasi algoritme terbaik diantaranya algoritme *Random Forest* dan *Support Vector Machine* (*SVM*) sebagai dasar penelitian selanjutnya untuk klasifikasi harapan hidup pasien penyakit hepatitis.

1.6 Manfaat

Adapun manfaat dari penelitian ini adalah klasifikasi algoritme terbaik dapat digunakan sebagai pembantu dalam meminimalisir keterlambatan proses diagnosis sehingga dapat menekan kenaikan angka penyakit hepatitis serta dapat mengurangi resiko angka kematian pada pasien. Selain itu hasil dari penelitian ini dapat dijadikan sebagai referensi untuk penelitian selanjutnya.