BAB I

PENDAHULUAN

1.2 Latar Belakang

Hati merupakan organ yang paling penting bagi manusia. Hati memiliki fungsi untuk pembentukan dan sekresi empedu, selain itu Hati sebagai alat penyimpanan glikogen, sintesa urea, metabolisme kolestrol dan lemak serta detoksifikan racun. Manusia tidak bisa hidup. Salah satu penyakit hati adalah *Hepatitis* atau *Liver*.

Hepatitis adalah penyakit peradangan pada hati, terjadi karena adanya toxin yang ada pada hati. Penyakit ini dapat menyerang semua orang tak terkecuali orang memiliki kekebalan tubuh yang tinggi. Penyakit *Hepatitis* atau *Liver* merupakan penyakit yang menular yang menjadi masalah kesehatan yang besar bagi masyarakat karena penularannya yang mudah. Peyakit *Hepatitis* atau *Liver* yang disebabkan oleh virus yang langsung menyerang pada organ hati. Penyakit ini dibagi menjadi lima jenis, yaitu, Hepatitis A, B, C, D, dan E (Kudesia & Wreghitt, 2009). Menurut WHO pada tahun 2017, jenis penyakit Hepatitis yang paling berbahaya adalah Hepatitis C(Kudesia & Wreghitt, 2009). Virus RNA adalah Virus memiliki RNA (asam ribonukleat) sebagai materi genetiknya. Asam nukleat biasanya untai tunggal RNA (ssRNA) tetapi mungkin juga untai ganda (dsRNA). Pada penyakit Hepatitis C memiliki tiga tingkatan yaitu sedang, *Fibrosis* atau berat, dan *Cirrhosis* atau kronis. Penularan Hepatitis C hanya melalui darah yang terkontaminasi hepatitis C Virus (HCV) atau berhubungan badan dengan orang yang terinfeksi. Pada tahun 2014 diperkirakan terdapat 28 juta penduduk Indonesia terserang penyakit hepatitis dan 14 juta diantaranya berpotensi untuk menjadi kronis dan dari yang kronis tersebut berpotensi menjadi kanker hati.Maka diperlukan penanganan yang khusus untuk penyakit Hepatitis C, dengan melakukan diagnosa dini terhadap penyakit Hepatitis C. Cara untuk melakukan diagnosa dini terhadap penyakit Hepatitis C adalah memanfaatkan teknologi informasi. Salah satunya menggunakan metode *data mining.*

Data mining yaitu pemrosesan suatu data yang menggunakan cara statistik, matematik, kecerdaan tiruan dan *machine learning* untuk mengidentifikasi

informasi pengetahuan potensial dan berguna yang tersimpan dalam database besar (Riadi, 2017).

Dalam klasifikasi *data mining* banyak algoritma yang dapat diterapkan untuk mengklasifikasi suatu data. Beberapa algoritmanya yaitu *Gaussian Naive Bayes* dan *K-Nearest Neighbor* (KNN). Penerapan algoritma *Gaussian Naive Bayes* dan *K-Nearest Neighbor* (KNN) sebagai proses klasifikasi penyakit Hepatitis C yang bertujuan untuk mendiagnosa dengan cara mengklasifikasi gejala-gejala yang dialami oleh pasien dengan mendapat nilai akurasi tertinggi. Oleh karena itu, penulis memutuskan menggunakan algoritma *Gaussian Naive Bayes* dan *K-Nearest Neighbor* (KNN), penulis memutuskan memilih algoritma *Gaussian Naive Bayes* dan *K-Nearest Neighbor* (KNN) dalam proses klasifikasi untuk melakukan perbandingan dan menemukan algoritma yang memiliki tingkat akurasi lebih tinggi dalam proses klasifikasi. Pada penelitian sebelumnya yang dilakukan oleh (Eka Wulansari Fridayanthie, 2017) yang berjudul Analisa Data Mining Untuk Prediksi Penyakit Hepatitis dengan Menggunakan Metode *Naive Bayes* dan *Support Vector Machine*, pada penelitian ini di dapat pengujian dengan menggunakan *Support Vector Machine* di dapat nilai akurasi adalah 68.38% sedangkan pengujian dengan menggunakan *Support Vector Machine* berbasis *Naive Bayes*di dapatkan nilai akurasi 83.71%. Pada penelitian selanjutnya yang dilakukan oleh (Prahudaya & Harjoko, 2017) yang berjudul Metode Klasifikasi Mutu Jambu Biji Menggunakan KNN Berdasarkan Fitur Warna dan Tekstur. Pada penelitian ini tingkat akurasi 91,25% pada nilai ketetanggan terbaik k-3. Pada penelitian yang dilakukan (Kamel et al., 2019) yang berjudul*Cancer Classification Using Gaussian Naive Bayes* Algorithm,dari penelitian tersebut didapatkan akurasi *Breast cancer* sebesar 98% dan *Lung cancer* sebesar 90%. Pada penelitian-penelitian sebelumnya yang sudah disebutkan berbeda dengan penilitian yang akan dilakukan penulis saat ini, dapat dilihat dari data yang digunakan, atribut yang digunakan, dan juga pada tujuan penelitian.

1.2 Rumusan Masalah

Adapun rumusan masalah yang dibahas dalam penelitian tugas akhir ini yaitu berapa hasil perbandingan nilai akurasi, presisi, dan recall pada algoritma *Gaussian Naive Bayes* dan *K-Nearest Neighbor* (KNN) untuk mendiagnosa penyakit hepatitis C ?

1.3 Batasan Masalah

Batasan masalah dalam penelitian tugas akhir ini sebagai berikut.

1. Data yang digunakanmerupakan dataset Hepatitis C pasien di Jerman pada tahun 2020
2. Atribut yang digunakan dalam proses klasifikasi berjumlah 13 atribut yaitu, Age, Sex, ALB, ALP, ALT, AST, BIL, CHE, CHOL, CREA, GGT, PROT, Category
3. Data yang digunakan dalam penelitian berjumlah 608 data dan diambil dari situs *https://archive.ics.uci.edu/ml/datasets/HCV+data*.
4. Tools yang digunakan adalah *Jupyter Notebook.*

1.4 Tujuan Penelitian

Tujuan dari penelitian tugas akhir ini yaitu untuk mengetahui hasil perbandingan nilai akurasi, presisi, dan recall pada algoritma *Gaussian Naive Bayes* dan *K-Nearest Neighbor* (KNN) untuk mendiagnosa penyakit Hepatitis C.

1.5 Maanfaat Penelitian

Manfaat dari penelitian tugas akhir ini adalah :

1. Untuk mengetahui perbandingan nilai akurasi, presisi, dan recall pada algoritma *Gaussian Naive Bayes* dan *K-Nearest Neighbor* (KNN).
2. Bisa dijadikan bahan referensi untuk melakukan penelitian selanjutnya.