

# **BAB I**

## **PENDAHULUAN**

### **1.1 LATAR BELAKANG MASALAH**

Stroke tergolong jenis penyakit yang paling banyak diderita masyarakat Indonesia dan memiliki tingkat kematian yang tinggi, stroke berperan merusak ataupun mengganggu fungsi otak yang dimana aliran darah yang mengalir untuk otak manusia terganggu dan hal ini memungkinkan pasien yang terkena penyakit stroke otak akan mengalami kelumpuhan atau kematian [1]. Pada tahun 2019 organisasi kesehatan dunia (WHO) mendudukkan stroke sebagai tujuh dari sepuluh penyebab utama kematian. Kementerian Kesehatan menggolongkan stroke sebagai penyakit katastrofik karena dampaknya luas secara ekonomi dan sosial. [2]. Stroke pada otak tidak bisa dianggap penyakit biasa karena jika dilihat dari gejala dan akibat dari stroke otak, penyakit ini harus dianggap sebagai penyakit serius.

Maka dari itu pentingnya kita untuk mengetahui informasi mengenai gejala ataupun penyebab dari penyakit stroke otak dengan salah satu teknik data mining yaitu klasifikasi dan juga dapat memprediksi mengenai penyakit stroke otak itu sendiri. Pada penelitian ini penulis akan menggunakan *Algoritma C4.5* karena *Algoritma C4.5* sangat baik untuk digunakan memecahkan kasus pada suatu masalah yang melibatkan teknik data mining yaitu klasifikasi.

Terdapat juga peneliti sebelumnya yang melakukan penelitian mengenai penyakit stroke, seperti pada penelitian [1] dengan menggunakan *Algoritma K-*

*Nearest Neighbor* (KNN) mendapatkan hasil 95%, lalu terdapat juga penelitian [3] yang dimana menggunakan *Algoritma C4.5* dan *Algoritma CART* dengan hold-out data 60:40 algoritma C4.5 memiliki tingkat akurasi yang baik mencapai 96% dan algoritma CART dengan akurasi hanya 95.76%, dan dengan hold-out data 70:30. Algoritma C4.5 mendapatkan hasil akurasi sebesar 95.76% sedangkan algoritma CART sebesar 95,11%, dan terdapat juga penelitian lain seperti [4] dengan menggunakan *Algoritma Naive Bayes* dan *Algoritma Decision Tree (j48)* dan mendapati nilai akurasi pada Algoritma Naive Bayes adalah 81.25% dan pada *Algoritma Decision Tree (J48)* mendapati hasil akurasi 87.5%.

Seperti yang bisa dilihat dari penelitian sebelumnya bahwa ketiga algoritma yang digunakan memiliki nilai akurasi yang sangat baik dalam mengklasifikasi dan memprediksi penyakit stroke otak, dan tentunya hal ini dapat menghasilkan informasi baru yang baik mengenai penyakit stroke otak itu sendiri dan dapat melakukan langkah lebih awal dalam mencegah penyakit stroke otak.

Sehubungan dengan permasalahan diatas dan pentingnya informasi mengenai penyakit stroke otak, penulis memutuskan untuk melakukan penelitian yang berjudul “*Penerapan Data Mining Dengan Algoritma C4.5 Untuk Memprediksi Penyakit Stroke Otak*”.

## **1.2 RUMUSAN MASALAH**

Berdasarkan dari latar belakang yang dijelaskan sebelumnya, maka rumusan masalah pada penelitian ini adalah :

1. Bagaimana cara melakukan prediksi pada data pasien mengenai penyakit Stroke otak dengan menerapkan *Algoritma C4.5*?
2. Bagaimana hasil pohon keputusan yang diperoleh *Algoritma C4.5* pada prediksi penyakit stroke otak?
3. Seberapa besar tingkat akurasi yang diperoleh *Algoritma C4.5* pada hasil dari prediksi penyakit Stroke Otak dengan menggunakan 2 jenis tools?
4. Bagaimana perbandingan performa dari kedua *tools* yang digunakan yaitu *WEKA* dan *RapidMiner*?

### 1.3 BATASAN MASALAH

Supaya pembahasan masalah tidak keluar dari tujuan penulis, maka batasan masalah yang terdapat pada penelitian ini adalah :

1. Data yang digunakan pada penelitian ini merupakan dataset dari bidang kesehatan yaitu data mengenai penyakit Stroke otak dan total keseluruhan data berjumlah 4981 data yang diperoleh dari website kaggle.com (<https://www.kaggle.com/datasets/jillanisofttech/brain-stroke-dataset?resource=download>).
2. Variabel atau Atribut yang terdapat pada dataset dan digunakan dalam penelitian ini yaitu *Gender, Age, Hypertension, Heart Disease, Ever Married, Work Type, Residence Type, AVG Glucose Level, BMI, Smoking Status*.
3. Aktivitas pengujian data pada penelitian ini akan menggunakan 2 *tools* yang berbeda yaitu *WEKA* dan *RapidMiner*.

## **1.4 TUJUAN DAN MANFAAT PENELITIAN**

### **1.4.1 Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah :

1. Untuk melakukan prediksi pada data mengenai penyakit Stroke Otak.
2. Untuk mendapatkan hasil akurasi yang optimal dengan menggunakan *Algoritma C4.5*.
3. Untuk menghasilkan sebuah pohon keputusan yang berguna untuk memudahkan penentuan gejala dari penyakit stroke otak.
4. Untuk mengetahui hasil akurasi yang diperoleh dari kedua *tools* yang digunakan yaitu *WEKA* dan *RapidMiner*.

### **1.4.2 Manfaat Penelitian**

Adapun manfaat dari penelitian ini adalah :

1. Memberikan informasi mengenai gejala dan penyakit Stroke agar mampu mengetahui penyebab dan akibat dari penyakit Stroke sehingga mampu mengambil langkah pencegahan lebih awal.
2. Dapat mengetahui seberapa besar tingkat akurasi *Algoritma C4.5* untuk memprediksi penyakit Stroke otak.
3. Mendapatkan hasil dari pohon keputusan yang diperoleh dari 2 *tools* yang berbeda dan dapat dipahami.
4. Dapat mengetahui informasi mengenai performa dari kedua *tools* yaitu *WEKA* dan *RapidMiner* dengan melihat perbandingan hasil akurasi yang diperoleh.

## **1.5 SISTEMATIKA PENULISAN**

Untuk memudahkan penyusunan dalam penulisan skripsi ini, penulis membagi beberapa bagian dari pembahasan kedalam beberapa bab. Adapun sistematika penulisan skripsi ini yaitu :

### **BAB I : PENDAHULUAN**

Pada bab ini berisi penjelasan mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan beserta manfaat penelitian, dan sistematika penulisan.

### **BAB II : LANDASAN TEORI**

Pada bab ini berisi kumpulan teori-teori dan pendapat dari para ahli yang berkaitan dengan topik permasalahan sesuai dengan permasalahan yang diteliti yang dimana pada penelitian ini berkaitan dengan salah satu teknik data mining.

### **BAB III : METODOLOGI PENELITIAN**

Pada bab ini berisi penjelasan mengenai kerangka kerja penelitian, dan metode-metode yang digunakan untuk melakukan penelitian ini, dan juga alat bahan yang akan dipakai.

### **BAB IV : ANALISIS**

Pada 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 dari kesimpulan yang berkaitan dengan tujuan penelitian dari hasil yang telah diperoleh, serta berisi saran yang berkaitan dengan hasil dari penelitian dan juga untuk peneliti selanjutnya.