

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Mengutip dari Endang Indriasih dkk (2018) berdasarkan hasil survei *Sample Registration System* (SRS) di Indonesia pada tahun 2014, di lihat dari beberapa data tentang penyakit yang mematikan di Indonesia penyakit stroke ini merupakan penyakit dengan tingkat resiko mematikan nomor satu di Indonesia dengan angka kematian sekitar 21,1% kasus kematian yang di akibatkan oleh penyakit dalam satu tahun terakhir, maka dari itu diperlukan suatu penanganan atau tindakan pendeteksian secara dini agar resiko penyakit stroke bisa di atasi secara cepat dan dapat mengurangi tingkat resiko yang lebih parah. salah satu cara yang dapat di gunakan untuk pendeteksian penyakit stroke yaitu teknik klasifikasi data mining. Stroke merupakan salah satu penyakit dengan tingkat kematian yang cukup tinggi di dunia maupun di Indonesia. Penyakit ini juga dapat menjadi sangat berbahaya jika tidak segera cepat ditangani karena penyakit ini juga dapat menyebabkan cacat permanen bahkan hingga menyebabkan kematian (Hutama dkk, 2018).

Pada penelitian sebelumnya yang dilakukan oleh Lishania dkk (2019) berjudul “Perbandingan Metode klasifikasi *Naive Bayes* Dan Metode *Decision Tree* Algoritma (J48) Pada Pasien Penyakit Stroke Di RSUD Abdul Wahab Sjahranie Samarinda “, dengan kesimpulan algoritma *naive bayes* mendapat akurasi terbesar yaitu sebesar 81,25%. Adelina dkk (2018) melakukan penelitian berjudul “Klasifikasi Tingkat Resiko Penyakit Stroke Menggunakan *GA-Fuzzy Tsukamoto*” dengan ini dapat di simpulkan *Fuzzy Tsukamoto* juga dapat diterapkan pada kasus pendiagnosisan risiko Stroke. Dari kedua penelitian tersebut dapat diketahui bahwa teknik data mining mampu untuk mengklasifikasikan penyakit stroke dengan baik.

Dalam data mining sendiri terdapat banyak metode atau algoritma dalam menangani kasus pengklasifikasian salah satunya yaitu KNN (*K-Nearest Neighbor*). berdasarkan penelitian yang di lakukan oleh Sulastri dkk (2019) yang berjudul “Analisa Perbandingan Klasifikasi Prediksi Penyakit Hepatitis Dengan

Menggunakan Algoritma *K-Nearest neighbor, Naive bayes dan Neural network*” Penelitian ini dilakukan dengan menggunakan data yang diambil dari UCI *Machine Learning* berupa data pasien hepatitis dengan Data berjumlah 155 record dan terdiri dari 19 variabel penjelas dan 1 variabel respon. menghasilkan kesimpulan bahwa algoritma KNN memiliki akurasi terbaik yaitu sebesar 93%. Dalam data mining terdapat juga teknik optimasi untuk meningkatkan kinerja algoritma menjadi lebih baik, salah satu teknik optimasi yaitu *Bagging*. Penelitian penelitian menggunakan teknik *Bagging* telah banyak di terapkan antara lain yaitu Rahmawati & Agustina (2020) yang berjudul “Implementasi Teknik *Bagging* Untuk Meningkatkan Kinerja J48 Dan *Logistic Regression* Dalam Prediksi minat Pembeli Online” dengan kesimpulan bahwa penggunaan teknik bagging terbukti mampu meningkatkan kinerja algoritma.

Berdasarkan uraian di atas, maka dalam penelitian ini akan dilakukan optimasi algoritma KNN dengan teknik *Bagging* dalam mengklasifikasikan penyakit stroke dengan judul “Klasifikasi Penyakit Stroke Menggunakan Algoritma KNN dengan optimasi *Bagging*”

### **1.2 Rumusan Masalah**

Berdasarkan dari latar belakang diatas maka rumusan masalah dari penelitian ini adalah sebagai berikut:

Berapakah tingkat dari akurasi presisi, dan *recall* pada klasifikasi penyakit stroke ini menggunakan algoritma *KNN* dengan pengoptimasian teknik *bagging*?

### **1.3 Batasan Masalah**

Diperlukan adanya suatu batasan masalah supaya tidak keluar dari pokok-pokok permasalahan yang telah di dirumuskan, maka ruang lingkup penelitian dibatasi pada:

1. *Dataset* yang digunakan ini merupakan dataset tentang penyakit stroke yang di ambil dari sumber situs resmi [www.Kaggle.com](http://www.Kaggle.com), yang telah diakses pada tanggal 2 agustus 2021 yang berjumlah sebanyak 4894 *data* dan 11 atribut.

2. Untuk *Tools* yang digunakan pada penelitian ini adalah *jupyter notebook*, dengan bahasa pemrograman python..
3. partisi data meliputi 80% data latih dan 20% data uji.

#### **1.4 Tujuan Penelitian**

Penelitian ini bertujuan untuk menerapkan algoritma KNN (K- Nearest Neighbor )

dan membanding dengan hasil dari optimasi bagging untuk mencari tingkat Akurasi, Presisi dan Recall

#### **1.5 Manfaat Penelitian**

Manfaat penelitian dari metode ini adalah sebagai berikut:

1. Bagi Penulis:

Hasil dari suatu penelitian ini dapat bermanfaat untuk meningkatkan wawasan bagi penulis sendiri dalam hal bidang data mining.

2. Bagi Peneliti Lain:

bagi peneliti yang lain dapat bermanfaat untuk dijadikan referensi bagi penelitian berikutnya khususnya dibidang data mining, dan diharapkan mampu menjadi referensi dalam penanganan penyakit stroke di kemudian hari.