

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi sangat penting dalam proses kehidupan manusia. Seiring dengan pertumbuhan teknologi informasi, membuka peluang terhadap pertumbuhan data yang relatif banyak. Sebelum semua data memiliki nilai dan divalidasi, diperlukan cara untuk memproses, menyimpan, dan mengelola semua data tersebut. Salah satu cara untuk memfasilitasi hal tersebut dengan adanya teknologi basis data yang dikenal dengan *data mining*.

Data mining merupakan proses pengolahan data menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang berguna dan pengetahuan yang terkait dari berbagai basis data besar. Terdapat beberapa teknik yang dapat kita gunakan dalam *data mining* untuk menemukan sebuah informasi baru, salah satu teknik dari *data mining* adalah klasifikasi.

Klasifikasi adalah metode pengelompokan data secara sistematis menurut aturan dan kaidah yang telah ditetapkan sebelumnya. Klasifikasi juga dapat diartikan sebagai pengelompokan data atau objek baru yang kelasnya belum diketahui atau kategorinya. Klasifikasi adalah salah satu metode statistik untuk pengelompokan atau pengkategorian data yang tersusun secara sistematis.

Ada beberapa algoritma yang biasa digunakan untuk klasifikasi data, antara lain *K-Nearest Neighbor*, *Decision Tree*, *Neighbor Weighted K-Nearest Neighbor* (NWKNN) dan algoritma Genetika. Banyak peneliti menerapkan algoritma *K-Nearest Neighbor* dalam pengklasifikasian, diantaranya yang dilakukan oleh (Atthalla et al., 2018) mengenai klasifikasi penyakit kanker payudara menggunakan metode *K-Nearest Neighbor*. Pada penelitian ini penghitungan jarak kemiripan menggunakan jarak *minkowski*, Jarak *minkowski* adalah jarak di dalam ruang vektor yang telah ditentukan yang bisa dianggap sebagai generalisasi dari kedua jarak *Euclidean* dan jarak *Manhattan*. Dengan

metode KNN diperoleh hasil paling akurat sebesar 93%. Kemudian juga mengenai penelitian yang pernah dilakukan oleh (Sari & Mahmudy, 2019) yang berjudul Penerapan Metode *Decesion Tree* dan Algoritma Genetika untuk Klasifikasi Risiko Hipertensi. Pada peneltian ini menghasilkan nilai akurasi pengujian algoritma DT-GA rata-rata akurasi tertinggi 84%. Kemudian pada penelitian yang dilakukan oleh (Yudha et al., 2018) yang berjudul klasifikasi risiko Hipertensi menggunakan metode *Neighbor Weighted K-Nearest Neighbor* (NWKNN). Hasil dari penelitian ini menunjukkan saat data latih yang digunakan sebanyak 100 data dan data uji sebanyak 25 data, nilai $K=10$, dan nilai $E=4$ dengan tingkat akurasi mencapai 88%. Tiga perbandingan tersebut menentukan hasil keputusan dari metode *K-Nearest Neighbor* lebih akurat dibandingkan dengan hasil keputusan metode *Decision Tree*, Algoritma Genetika dan *Neighbor Weighted K-Nearest Neighbor* (NWKNN), sehingga dari itu peneliti mengambil metode algoritma *K-Nearest Neighbor*.

Metode yang digunakan dalam klasifikasi data adalah algoritma *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* (KNN) adalah algoritma yang sangat sederhana dengan cara mengelompokkan data baru dengan K tetangga terdekat. (Putra et al., 2019)

Hipertensi adalah peningkatan tekanan darah sistolik lebih besar atau sama dengan 140 mmHg, dan peningkatan tekanan diastolik lebih besar atau sama dengan 90 mmHg. Hipertensi juga sering disebut *Silent Killer* karena tidak semua penderita Hipertensi menyadari penyakit yang dideritanya. Hipertensi adalah penyebab utama terjadinya gagal jantung, stroke, dan gagal ginjal. Banyak faktor yang dapat mempengaruhi Hipertensi yaitu faktor yang dapat dikendalikan atau tidak dikendalikan. Berdasarkan data nasional hasil Riskesdas 2018 menunjukkan bahwa prevalensi penduduk dengan tekanan darah tinggi sebesar 34,1%. Prevalensi Hipertensi di provinsi Aceh 26,45% dan prevalensi di Aceh Tenggara sebesar 24,75%.

Berdasarkan latar belakang masalah diatas penulis tertarik membuat judul **“Penerapan Metode *K-Nearest Neighbor* untuk Klasifikasi Penyakit**

Hipertensi”. Dengan tujuan memudahkan masyarakat dalam mengetahui berbagai jenis Hipertensi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka permasalahan yang penulis rumuskan dalam penelitian ini, adalah:

1. Bagaimana merancang sistem klasifikasi penyakit Hipertensi di RSUD H. Sahudin.
2. Bagaimana menerapkan algoritma *K-Nearest Neighbor* dalam mengklasifikasi jenis penyakit Hipertensi.

1.3 Batasan Masalah

Agar fokus penelitian ini tetap terjaga maka batasan masalah penelitian tugas akhir ini adalah:

1. Data yang digunakan adalah data hasil rekam medis dan uji laboratorium RSUD H. Sahudin tahun 2021 dan 2022 dengan jumlah data 478.
2. Atributnya yaitu usia, tensi sistolik, tensi diastolik, asam urat, glukosa, kolestrol, berat badan, tinggi badan dan BMI.
3. Bahasa pemrograman yang digunakan yaitu *pyhton* dan menggunakan database *MySQL*.
4. Metode yang digunakan algoritma *K-Nearest Neighbor*.
5. Hasil *output* nya berupa Normal, Pre-Hipertensi, Hipertensi 1 dan Hipertensi 2.

1.4 Tujuan Penelitian

Tujuan penelitian pada tugas akhir ini adalah untuk:

1. Merancang sistem klasifikasi penyakit Hipertensi di RSUD H. Sahudin.
2. Menerapkan algoritma *K-Nearest Neighbor* dalam mengklasifikasi jenis penyakit Hipertensi.

1.5 Manfaat Penelitian

Manfaat yang akan diperoleh dari hasil penelitian ini adalah:

1. Dengan adanya sistem ini, memudahkan masyarakat dalam mengklasifikasikan penyakit Hipertensi.
2. Memudahkan masyarakat mengklasifikasikan secara efisien dan efektif.
3. Membantu tenaga medis untuk meningkatkan pelayanan kepada masyarakat tentang kondisi penyakit Hipertensi tersebut.
4. Sistem dapat digunakan oleh masyarakat untuk mengetahui kondisi kesehatan khususnya Hipertensi.