

BAB I

PENDAHULUAN

1.1. Latar Belakang

Saat ini kebutuhan terhadap analisis suatu data sangat dibutuhkan. Pesatnya perkembangan data yang semakin tinggi mendorong untuk memanfaatkan data dalam penggalian informasi maupun pengetahuan. Salah satu pendekatan yang dapat dilakukan untuk menganalisis sekumpulan data yaitu dengan mengklasifikasi data. Klasifikasi termasuk dalam salah satu teknik data mining yang digunakan untuk membangun suatu model dari sampel data yang belum terklasifikasi untuk digunakan mengklasifikasi sampel data baru ke dalam kelas-kelas yang sejenis (Sartika & Sensuse, 2017).

Menurut Prasetyo dalam (Arisandy, 2017) klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam metode klasifikasi memiliki dua proses yang dapat dilakukan yaitu dengan membangun model untuk disimpan sebagai memori dan menggunakan model tersebut untuk melakukan pengenalan atau klasifikasi atau prediksi pada suatu data lain supaya diketahui di kelas mana objek data tersebut dimasukkan berdasarkan model yang telah disimpan dalam memori.

Beberapa metode klasifikasi yang umum digunakan yaitu seperti *Artificial Neural Network* (ANN), *Support Vector Machines* (SVM), *Decision Tree*, *Bayesian*, dan sebagainya. Selain beberapa metode tersebut, terdapat metode yang memiliki tingkat akurasi tinggi pada saat proses klasifikasi yaitu *Naïve Bayes*. Menurut Han dan Kamber dalam (Arisandy, 2017), klasifikasi *Naïve Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam basis data dengan jumlah yang besar. Penelitian yang dilakukan oleh Heru Sulistiono (2015) berjudul “*Kajian Penerapan Algoritma c4.5, Neural Network Dan Naïve Bayes Untuk Klasifikasi Mahasiswa Yang Bermasalah Dalam Registrasi*”. Algoritma *Naive Bayes* memiliki tingkat akurasi yang paling tinggi dibandingkan dengan algoritma *Neural Network* dan *C4.5* dengan nilai persentase akurasi 93,58%.

Selain *Naive Bayes*, algoritma *K-Nearest Neighbor* juga memiliki tingkat akurasi yang tinggi. Penelitian yang dilakukan oleh Mustakim (2016) berjudul “*Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa*” tingkat akurasi dari algoritma *K-Nearest Neighbor* sebesar 82%.

Dari kedua penelitian yang disebutkan diatas menunjukkan bahwa algoritma *Naive Bayes* dan *K-Nearest Neighbor* memiliki tingkat akurasi yang tinggi untuk mengklasifikasi data. Oleh karena itu peneliti ingin melakukan sebuah penelitian dengan membandingkan algoritma *Naive Bayes* dan algoritma *K-Nearest Neighbor* menggunakan data diagnosa penyakit jantung yang diambil dari penyedia layanan dataset publik *UCI Machine Learning*.

Pada tahun 2005 sedikitnya 17,5 juta atau setara dengan 30% kematian di seluruh dunia disebabkan oleh penyakit jantung. Menurut Badan Kesehatan Dunia (WHO) dalam (Oemiati & Rustika, 2015), 60% dari seluruh penyebab kematian penyakit jantung adalah penyakit jantung koroner (PJK). Faktor gejala yang terdiagnosa sebagai penyakit jantung antara lain adalah jenis sakit dada (*chest pain*), tekanan darah tinggi (*tresbps*), kolesterol (*chol*), nilai tes EKG (*resting electrodiagraphic "restacg"*), denyut jantung (*thalach*) dan kadar gula (*fasting blood sugar "FBS"*), dan beberapa faktor lainnya yang mengidentifikasi bahwa seseorang mempunyai penyakit jantung (Rifai, 2013).

Berdasarkan pembahasan diatas peneliti mengambil judul “*Perbandingan Algoritma Naive Bayes dan Algoritma K-Nearest Neighbor untuk Klasifikasi Penyakit Jantung*”.

1.2. Rumusan Masalah

Permasalahan yang dijadikan objek dalam penelitian tugas akhir ini berdasarkan latar belakang yang telah dipaparkan adalah:

1. Berapa hasil perbandingan akurasi dan presisi dari algoritma *Naive Bayes* dan algoritma *K-Nearest Neighbor* untuk klasifikasi penyakit jantung.
2. Berapa nilai *K* pada algoritma *K-Nearest Neighbor* yang menghasilkan tingkat akurasi dan presisi terbaik untuk klasifikasi penyakit jantung.

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian tugas akhir ini adalah:

1. Menganalisa perbandingan akurasi, dan presisi pada algoritma *Naive Bayes* dan algoritma *K-Nearest Neighbor* untuk klasifikasi penyakit jantung.
2. Menganalisa jumlah nilai *K* terbaik untuk menghasilkan akurasi, dan presisi pada algoritma *K-Nearest Neighbor* untuk klasifikasi penyakit jantung.

1.4. Batasan Masalah

Untuk lebih memfokuskan pengerjaan penelitian ditetapkan pembahasan hanya dibatasi pada:

1. Data yang digunakan pada penelitian ini yaitu data diagnosa penyakit jantung yang disediakan oleh (UCI) Machine Learning University sebanyak 300 *record data* pasien yang diperiksa dan sebanyak 163 pasien terdeteksi sehat dan 137 terdeteksi sakit jantung.
2. Atribut yang digunakan terdiri dari 12 atribut, yaitu umur, jenis kelamin, sakit dada, tekanan darah, kolesterol, kadar gula, elektrokardiografi, tekanan jantung, angina induksi, *oldpeak*, *slope*, dan *thallium*.
3. Hasil akhir pengujian berupa informasi perbandingan akurasi, dan presisi.
4. Pengujian dilakukan dengan bantuan Aplikasi *RapidMiner Studio*.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberi manfaat bagi peneliti, mahasiswa, dan bagi penelitian lain.

1. Bagi Peneliti

Adapun manfaat dalam penelitian ini yaitu untuk mengetahui perbandingan akurasi algoritma *Naive Bayes* dan *K-Nearest Neighbor* dalam proses klasifikasi penyakit jantung.

2. Bagi Mahasiswa

Penelitian ini diharapkan dapat membantu mahasiswa dalam memperluas wawasan mengenai algoritma *Naive Bayes* dan *K-Nearest Neighbor*.

3. Bagi Penelitian Lain

Penelitian ini diharapkan dapat dijadikan sebagai referensi bagi penelitian lain yang ingin menggunakan algoritma *Naive Bayes* dan *K-Nearest Neighbor* dalam mengklasifikasi data maupun memprediksi data, begitu pun juga dapat digunakan sebagai referensi untuk mengembangkan penelitian lebih lanjut yang berkaitan dengan algoritma *Naive Bayes* dan *K-Nearest Neighbor*.

