

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi terdapat dua proses yang 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 (Prasetyo, 2012). Algoritma Klasifikasi yang umum digunakan antara lain *Artificial Neural Network (ANN)*, *Support Vector Machines (SVM)*, *Decision Tree*, Bayesian, dan sebagainya (Prasetyo, 2012).

Dari beberapa metode klasifikasi tersebut, terdapat metode yang memiliki tingkat akurasi tinggi yaitu *Naïve Bayes*. Hal tersebut diungkapkan oleh Han and Kamber (2006) klasifikasi *Naïve Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam basis data dengan jumlah yang besar. Penelitian yang dilakukan oleh Paulus Dian Wicaksana (2015) berjudul “*Perbandingan Algoritma K-Nearest Neighbors dan Naive Bayes Untuk Studi Data Wisconsin Diagnosis Breast Cancer*” Naive bayes menghasilkan akurasi 97.5% dan lebih tinggi dari K-Nearest Neighbor 95,6%.

Selain *Naïve Bayes*, algoritma C4.5 juga memiliki tingkat akurasi yang tinggi. Penelitian yang dilakukan oleh Puji Astuti (2016) berjudul “*Komparasi Penerapan Algoritma C45, Knn Dan Neural Network Dalam Proses Kelayakan Penerimaan Kredit Kendaraan Bermotor*” akurasi tertinggi yang dihasilkan C4.5 yaitu 92,89% dibandingkan K-nn 77,78% dan Nn 91,1%.

Dari kedua penelitian diatas menunjukkan bawa algoritma Naive Bayes dan algoritma C4.5 memiliki tingkat akurasi yang tinggi pada proses klasifikasi data.

Namun Bagaimana jika kedua algoritma tersebut dibandingkan dengan menggunakan data yang sama. Bagaimana akurasi yang akan dihasilkan? Apakah algoritma Naive Bayes lebih unggul dengan jumlah data yang banyak dibandingkan algoritma C4.5? atau sebaliknya?

Oleh karena itu peneliti ingin melakukan sebuah penelitian dengan membandingkan algoritma Naive Bayes dan algoritma C4.5 menggunakan multi data yang peneliti ambil dari penyedia layanan dataset publik, dengan kriteria masing-masing data yang berbeda, hal ini bertujuan untuk menganalisa hasil akurasi dari kedua metode tersebut terhadap data yang diujikan.

Berdasarkan pembahasan diatas peneliti mengambil judul “*Analisis Perbandingan Algoritma Naive Bayes dan Algoritma C4.5 Untuk Klasifikasi Multi Data*”.

1.2 Rumusan masalah

Bedasarkan latar belakang yang dipaparkan di atas dapat diambil rumusan masalah :

1. Bagaimana akurasi algoritma Naive Bayes dan algoritma C4.5 dalam klasifikasi multi data?
2. Bagaimana analisa akurasi yang dihasilkan oleh algoritma Naive Bayes dan algoritma C4.5 dalam klasifikasi multi data?

1.3 Tujuan

Adapun tujuannya dari penelitian ini yaitu :

1. Menghitung tingkat akurasi algoritma Naive Bayes dan algoritma C4.5 dalam klasifikasi multi data.
2. Menganalisa akurasi yang dihasilkan oleh algoritma Naive Bayes dan algoritma C4.5 dalam klasifikasi multi data.

1.4 Batasan masalah

Agar pembahasan masalah tidak menyimpang dari tujuan, maka berikut adalah beberapa batasan yang perlu di buat, yaitu:

1. Membahas tentang perbandingan algoritma Naive Bayes dan algoritma C4.5
2. Tingkat akurasi dan Time taken dari masing-masing algoritma yang dijadikan sebagai pembanding.
3. Data yang digunakan pada penelitian ini adalah 10 dataset yang diambil dari UCI (University of California, Irvine) yaitu, (1) *Balance Scale Weight & Distance*, (2) *Car Evaluation*, (3) *Haberman's Survival Data*, (4) *Iris Plants*, (5) *Qualitative_Bankruptcy*, (6) *Tic-Tac-Toe Endgame*, (7) *lenses*, (8) *Contraceptive Method Choice*, (9) *Adult*, (10) *Poker*.
4. Pengujian dilakukan dengan *Cross Validation* dan *Use training set*.
5. Pengujian dilakukan dengan bantuan Aplikasi *Weka*.

1.5 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat bagi Peneliti, Mahasiswa dan bagi penelitian lain.

1) Bagi Peneliti

Adapaun manfaat dalam penelitian ini yaitu untuk mengetahui bagaimana akurasi yang dihasilkan dari algoritma *Naïve Bayes* dan algoritma *C4.5*.

2) Bagi Mahasiswa.

Penelitian ini diharapkan dapat membantu mahasiswa dalam memperluas wawasan mengenai algoritma *Naïve Bayes* dan algoritma *C4.5*.

3) Bagi Penelitian Lain

Penelitian ini diharapkan dapat dijadikan sebagai referensi bagi penelitian lain yang ingin menggunakan mengembangkan tentang perbandingan algoritma.