

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Klasifikasi merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan “jika-maka”, berupa *decision tree*, *formula matematis* atau *neural network*. Metode-metode klasifikasi antara lain C4.5, *RainForest*, *Naïve Bayes*, *neural network*, *genetic algorithm*, *fuzzy*, *case-based reasoning*, dan *k-Nearest Neighbor* (Arriawati A S, 2011).

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 Luki Dwi Arisandi (2017) pada penelitian ini peneliti memakai 10 data set dan 3 teknik pengujian K fold validation berjudul “*Perbandingan Algoritma Naïve Bayes dan C.45 Untuk klasifikasi multi data*” dimana peneliti ini C.45 lebih unggul dalam *Naïve bayes*, C.45 memiliki akurasi yang yaitu sebesar 89,46% sedangkan *Naïve Bayes* menghasilkan akurasi 79,39%.

Selain *Naïve Bayes*, *algoritma K-Nearest Neighbor* atau dapat disingkat dengan K-NN adalah salah satu metode non parametrik yang digunakan dalam pengklasifikasian. Pada penelitian sebelumnya yang dilakukan oleh Nur Khotimah dan Deden Istiawan (2018) berjudul “*Perbandingan Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour untuk Prediksi Lahan Kritis di Kabupaten Pematang*” Dari hasil perhitungan pengujian akurasi algoritma K-Nearest Neighbor mampu menghasilkan akurasi sebesar 73,91%.

Dari kedua penelitian tersebut menunjukkan bahwa algoritma Naïve Bayes dan algoritma K-Nearest Neighbor 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 Naïve Bayes lebih unggul dengan jumlah data yang banyak dibandingkan algoritma K-Nearest Neighbor, karena K-Nearest Neighbor atau Naive bayes merupakan algoritma data mining yang sama sama baik dalam mengklasifikasi dan regresi data.

Oleh karena itu peneliti ingin melakukan sebuah penelitian dengan membandingkan algoritma Naïve Bayes dan algoritma K-Nearest Neighbor menggunakan multi data yang peneliti ambil dari penyedia layanan dataset publik UCI (University of California, Irvine), dengan kriteria masing-masing data yang berbeda meliputi variabel, tipe data dan jumlah data, Hal ini bertujuan untuk menganalisa hasil akurasi dari kedua metode tersebut terhadap data yang diujikan. Berdasarkan pembahasan diatas peneliti mengambil judul “*Analisis Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Multi Data Set*”.

1.2 Rumusan masalah

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

1. Bagaimana penerapan algoritma Naïve Bayes dan K-Nearest Neighbor dalam klasifikasi multi data set?
2. Bagaimana akurasi algoritma Naïve Bayes dan K-Nearest Neighbor dalam klasifikasi multi data set?

1.3 Tujuan

Adapun Tujuan dari penelitian ini adalah

1. Menerapkan algoritma Naïve Bayes dan K-Nearest Neighbor dalam klasifikasi multi data set
2. Menghitung akurasi algoritma Naïve Bayes dan K-Nearest Neighbor dalam klasifikasi multi data set.

1.4 Batasan masalah

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

1. Tingkat akurasi dan algoritma yang dijadikan sebagai pembanding.
2. Data yang digunakan pada penelitian ini adalah 1 dataset yang diambil dari UCI (University of California, Irvine) yaitu, (1) *bank Marketing* berjumlah 45212 record (2) *tic-tac-toe endgame* berjumlah 958 record.
3. Pengujian dilakukan dengan *Use training set bank marketing* dengan 1 data training berjumlah 11 atribut yang dipakai .
4. Pengujian dilakukan dengan bantuan Aplikasi *Weka*.

1.5 Manfaat

Penelitian ini diharapkan dapat memberikan pemahaman tentang Analisa klasifikasi multi data set menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor lebih spesifik mengenai jumlah data , jenis data , variabel dan tipe data dan diharapkan dapat membantu peneliti berikutnya.