

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Zam-Zam Time merupakan Izin Usaha Mikro Kecil (IUMK) yang bergerak dibidang bisnis minuman. Zam-Zam Time saat ini memiliki banyak *outlet* yang tersebar di Purwokerto dan sekitarnya. Banyaknya *outlet* menandakan semakin banyak pelanggan yang membeli produk Zam-Zam Time maka pelayanan dan produk yang dijual harus memiliki kualitas baik sehingga dapat membantu tercapainya bisnis yang diinginkan. Tingkat kepuasan pelanggan bisa dilihat dari cara pelayanan dan penilaian pada produk itu sendiri dengan produk sesuai pada kebutuhan dan ekspektasi pelanggan. Zam-Zam Time memiliki jumlah produk sebanyak 8 varian rasa. Belum adanya penilaian pelayanan dan produk pada Zam-Zam Time membuat Zam-Zam Time kesulitan untuk mengevaluasi barista dan apakah produk Zam-Zam Time termasuk dalam golongan Laris atau Kurang Laris. Kepuasan pelanggan terhadap Zam-Zam Time merupakan salah satu faktor ketika Zam-Zam Time akan melakukan peningkatan mutu pelayanan dan menentukan bahwa produk Zam-Zam Time tergolong laris atau kurang laris. Berdasarkan masalah ini dibutuhkan sebuah teknik klasifikasi tingkat kepuasan pelanggan terhadap produk Zam -Zam Time. Produk Zam-Zam Time tergolong laris atau kurang laris.

Di era digitalisasi sekarang ini pengolahan data akan lebih mudah jika dilakukan dalam bidang data mining yang merupakan sebuah teknik menemukan korelasi yang bermakna dengan jumlah data besar yang disimpan pada repositori menggunakan teknologi statistika dan matematika. Metode dalam data mining dapat dikategorikan sebagai *supervised learning* ditandai dengan adanya variabel target yang terdiri dari estimasi, prediksi serta klasifikasi sedangkan untuk *unsupervised learning* ditandai dengan tidak ada variabel target

yang terdiri dari *clustering* dan asosiasi [1] . Metode klasifikasi akan diterapkan sebagai metode dalam penelitian ini karena dalam klasifikasi memiliki variabel target dimana proses klasifikasi data dapat dilakukan dengan pendekatan klasifikasi data mining yaitu mengelola data secara sistematis sehingga hasil klasifikasi dapat dengan mudah untuk dimengerti [2] . Klasifikasi merupakan suatu model data mining yang proses pengelompokannya secara sistematis akan dikonstruksi dalam kategori label menurut jenis dengan ciri yang sama, standar data yang sudah dikategorikan tersebut dapat divisualisasikan [3][4] . Atribut label yang digunakan bertipe kategorikal merupakan ciri data yang digunakan dalam klasifikasi. Data yang mengalami klasifikasi memiliki dua tahapan yaitu yang pertama dilakukan proses *Training* data dan *Testing* data yang kemudian dianalisa menggunakan algoritma klasifikasi [5] . Terdapat beberapa algoritma yang dapat digunakan dalam proses klasifikasi diantaranya yaitu: *K-Nearest Neighbor* (KNN) [6][7][8] yang merupakan sebuah proses klasifikasi yang memanfaatkan data terdekat dalam pengkajian sebagai data *Training* dengan entitas baru sebagai data *Testing* , *Support Vector Machine* (SVM) [9][10][11] adalah sebuah cara penelaahan untuk analisis dalam mengatasi dua kelas permasalahan untuk mendeteksi *hyperplane* dengan mengoptimalkan jarak kelas, *Naïve Bayes* [12][13][14] yaitu sebuah metode klasifikasi statistika yang terstruktur digunakan dalam memprediksi probabilitas suatu kelas berdasarkan distribusi data dalam dokumen klasifikasi statistik. C4.5 [15][16][17] merupakan suatu algoritma yang dimanfaatkan untuk membangun sebuah pohon keputusan serta digunakan dalam pembelajaran mesin dan penambangan data.

Beberapa penelitian sebelumnya [18][19][20][21][22] didapat algoritma terbaik dalam proses klasifikasi yaitu Algoritma *Naïve Bayes* dengan kelebihan mudah dipahami dan diimplementasikan, algoritma yang tergolong cepat, dapat menangani data numerik dan kategorikal. Perhitungan *Teorema Bayes* yang sederhana sehingga dapat mengurangi kerumitan perhitungan menjadi perkalian sederhana dari probabilitas. Waktu yang dibutuhkan lebih sedikit untuk data yang dilatih dengan estimasi parameter yang dibutuhkan dalam proses klasifikasi [23] metode yang simpel dan waktu komputasi yang singkat [24] kemudian Algoritma

C4.5 memiliki keunggulan tahan terhadap data *noise*, dapat menangani variabel diskrit, kontinu[25] serta hilang dan bisa memangkas cabang pohon keputusan yang lain, cabang yang memvisualisasikan hasil dari atribut yang *Testing* dalam sebuah kelas sehingga pohon keputusan dengan mudah diubah menjadi aturan klasifikasi dan dipetakan nilai atribut ke kelas baru, pohon keputusan yang mudah dimengerti, dibangun, dan tidak membutuhkan sampel data yang banyak seperti algoritma klasifikasi lainnya.

Berdasarkan kelebihan Algoritma C4.5 dan Naïve Bayes maka pada penelitian ini diperlukan sebuah komparasi dari kedua algoritma tersebut untuk menemukan algoritma baik yang terlihat pada nilai akurasi yang merupakan hasil dan waktu komputasi yang diperlukan masing-masing algoritma dalam proses pengklasifikasian tingkat kepuasan pelanggan terhadap produk Zam-Zam Time.

Adapun tujuan dari penelitian ini yaitu untuk mengukur dan menganalisis algoritma terbaik dengan melihat hasil nilai akurasi dan waktu komputasi antara Naïve Bayes dan C4.5 dalam klasifikasi produk Zam-Zam Time berdasarkan tingkat kepuasan pelanggan. Hasil penelitian ini dapat digunakan sebagai bahan evaluasi Zam-Zam Time sehingga dapat meningkatkan hasil omset.

## **1.2 Rumusan masalah**

Berdasarkan latar belakang diatas, Zam-Zam Time masih menggunakan sistem yang manual, belum adanya penilaian secara objektif dari pelanggan terhadap produk Zam-Zam Time, sehingga dalam proses pengklasifikasian produk tergolong Laris atau Kurang Laris membutuhkan waktu yang lama, oleh karena itu dibutuhkan sebuah teknik untuk menyelesaikan masalah tersebut dengan proses klasifikasi menggunakan algoritma. Analisis algoritma mana yang memiliki kinerja paling tinggi pada nilai akurasi dan waktu komputasi antara C4.5 dan Naïve bayes dalam klasifikasi produk Zam-Zam Time Laris atau Kurang Laris berdasarkan tingkat kepuasan pelanggan.

### **1.3 Pertanyaan Penelitian**

Menurut rumusan masalah di atas, seberapa besar kinerja menurut nilai akurasi dan waktu komputasi pada Algoritma C4.5 dan Naïve bayes untuk klasifikasi produk Zam-Zam Time tergolong Laris atau Kurang Laris berdasarkan tingkat kepuasan pelanggan ?

### **1.4 Batasan Masalah**

Dengan latar belakang tersebut maka dapat disebutkan batasan masalah dalam penelitian ini yaitu :

- a. Penelitian ini terbatas pada dataset yang bersifat privat atau primer diambil langsung dari hasil kuesioner tingkat kepuasan pelanggan pada produk Zam-Zam Time.
- b. Record yang digunakan adalah hasil kuesioner tingkat kepuasan pelanggan pada produk Zam-Zam Time.
- c. Klasifikasi Laris atau Kurang Laris berdasarkan pada nilai yang diberikan responden saat mengisi kuesioner yang telah disebar.
- d. Mengetahui faktor utama yang mempengaruhi Laris atau Kurang Laris dari hasil klasifikasi.
- e. Menganalisis nilai akurasi dan waktu komputasi yang dibutuhkan pada Algoritma Naïve Bayes dan C4.5 untuk klasifikasi produk Zam-Zam Time berdasarkan hasil kuesioner tingkat kepuasan pelanggan.

### **1.5 Tujuan**

Penelitian ini memiliki tujuan untuk mengukur dan menganalisis nilai akurasi serta waktu komputasi yang dihasilkan Algoritma C4.5 dan Naïve Bayes dalam proses klasifikasi produk Zam-Zam Time tergolong Laris atau Kurang Laris berdasarkan tingkat kepuasan pelanggan.

## **1.6 Manfaat**

Manfaat yang diharapkan pada penelitian ini adalah untuk mendapatkan Algoritma yang memiliki nilai akurasi dan waktu komputasi terbaik antara Naïve Bayes dan C4.5 pada proses klasifikasi produk Zam-Zam Time tergolong Laris atau Kurang Laris berdasarkan tingkat kepuasan pelanggan. Hasil klasifikasi nantinya dapat digunakan untuk bahan evaluasi Zam-Zam Time dalam meningkatkan mutu pelayanan dan kualitas produk.