

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

*Data Mining* adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (induction-based learning) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *Knowledge Discovery in Databases (KDD)* adalah penerapan metode saintifik pada data mining. Dalam konteks ini data mining merupakan satu langkah dari proses *KDD*. Salah satu metode yang digunakan dalam *Data Mining* adalah Algoritma *C4.5*.

Algoritma *C4.5*. merupakan kelompok algoritma pohon keputusan (decision tree). Algoritma ini mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Algoritma *C4.5* adalah salah satu metode untuk membuat decision tree berdasarkan training data yang telah disediakan. Algoritma *C4.5* merupakan pengembangan dari *ID3*. (Eka Pandu Cynthia, Edi Ismanto 2018)

Prediksi Loyalitas Pelanggan, yang menyatakan bahwa model klasifikasi pohon keputusan untuk memprediksi loyalitas pelanggan dan melihat variabel yang paling berpengaruh. Obyek dalam penelitian ini adalah pelanggan. Pembentukan model prediksi menggunakan metode C4.5 pada Algoritma C4.5 dilakukan perhitungan entropy dan information gain dimana atribut loyalitas pelanggan sebagai atribut tujuan, sedangkan harga, pelayanan, promosi, citra perusahaan dan kepercayaan sebagai atribut sumber untuk memperoleh node akar dan node lainnya. Berdasarkan hasil klasifikasi menggunakan Algoritma C4.5 menunjukkan bahwa diperoleh akurasi mencapai 97.5% yang menunjukkan bahwa Algoritma C4.5 cocok digunakan untuk menghitung prediksi penjualan barang (Kiki Rosita Dewi, dkk, 2020)

Pada Penelitian sebelumnya tentang Komparasi Metode Klasifikasi *Data mining* Algoritma C4.5 dan *Naive bayes* Untuk Prediksi Penyakit Hepatitis. Pada penelitian ini menghasilkan klasifikasi data mining Algoritma C4.5 menghasilkan akurasi 77,29% dan nilai AUC 0,846 yang termasuk dalam *Good Classification*. *Naive bayes* menghasilkan akurasi 83,71% dan nilai AUC 0,812. Dengan demikian dapat disimpulkan bahwa kedua metode ini akurat dalam melakukan prediksi untuk penyakit hepatitis. Algoritma C4.5 lebih unggul dari Naive bayes karena memiliki nilai AUC 0,846 dengan kategori *Good Clasification* (Ade Purwantoa, dkk, 2020)

Penelitian terdahulu *Algoritma C4.5* dapat membantu untuk mengambil keputusan penerimaan mahasiswa baru yang dapat lulus tepat waktu dan mahasiswa yang lulus terlambat dapat terprediksi lebih awal. *Algoritma C4.5* digunakan pada penelitian ini untuk memprediksi penerimaan peserta didik baru pada SMK Hasanah. Algoritma ini menganalisa set data sampel dan membangun sebuah *classifier* yang

harus mempunyai sebuah kapasitas untuk mengatur kasus latihan dan kasus percobaan dengan akurat (Riki Winanjaya, dkk, 2019)

Menurut Cambridge International Dictionaries pelanggan adalah a person who buys goods or a services (seseorang yang membeli suatu barang atau jasa). Kata kepuasan (satisfaction) berasal dari bahasa Latin “satis” (artinya cukup baik, memadai) dan “facio” (melakukan atau membuat). Menurut Richard Oliver (dalam Zeithaml., dkk 2018) kepuasan pelanggan adalah respon pemenuhan dari pelanggan terhadap suatu produk atau layanan itu sendiri yang telah memenuhi kebutuhan dan harapan pelanggan.

Implementasi *Data Mining* dengan metode *Algoritma C4.5* diatas memiliki kesamaan dengan masalah yang akan penulis bahas. Penulis mengangkat judul penelitian tentang **“KLASIFIKASI PELANGGAN AKTIF DALAM MENINGKATKAN PENJUALAN DENGAN MENGGUNAKAN ALGORITMA C4.5”**

**a. Perumusan Masalah**

Berdasarkan latar belakang di atas adapun yang menjadi pokok permasalahan dalam penelitian ini adalah :

1. Bagaimana cara menentukan kriteria yang menjadi penilaian pembelian terhadap pelanggan dengan menggunakan Algoritma C4.5?
2. Bagaimana cara menerapkan algoritma C4.5 dalam menganalisa jumlah pelanggan aktif?

**b. Batasan Masalah**

Agar penelitian ini lebih terarah dan dapat mencapai suatu sasaran serta menghindari terjadinya penyimpangan dari penelitian, maka penulis memberikan ruang lingkup permasalahan yang akan dibahas. Dalam penelitian ini penulis hanya membahas hal-hal sebagai berikut :

1. Data yang digunakan adalah data pelanggan pada PT. MULTI DAYA PRIMA.
2. Penelitian ini membahas tentang analisa jumlah pelanggan aktif dengan Algoritma C4.5.

**c. Tujuan Penelitian**

Tujuan penelitian antara lain sebagai berikut :

1. Untuk mengetahui kriteria apa saja yang menjadi penilaian terhadap pelanggan.
2. Mengetahui bagaimana proses data mining dapat membantu dalam menganalisa jumlah pelanggan aktif pada PT. MULTI DAYA PRIMA.
3. Menganalisa data pelanggan untuk mengetahui pelanggan aktif dan tidak aktif.
4. Merancang setiap *rule* dengan membuat pohon keputusan.
5. Menguji *rule* yang dibangun sesuai dengan data yang ada sehingga akan didapatkan hasil yang diinginkan.

d. **Manfaat Penelitian**

Manfaat yang diperoleh dari penelitian ini adalah:

1. Dapat memberikan informasi kepada perusahaan mengenai jumlah pelanggan aktif yang dimiliki oleh perusahaan.
2. Dengan mengetahui jumlah pelanggan aktif, diharapkan pihak perusahaan dapat menambah kualitas pelayanan, sehingga pelanggan pada perusahaan akan bertambah.

e. **Sistematika Penulisan**

Sistematika penulisan dalam pembuatan laporan penelitian adalah sebagai berikut :

**1. Bab I Pendahuluan**

Bab ini meliputi Latar Belakang, Perumusan Masalah, Batasan Masalah, Tujuan Penelitian, Sistematika Penulisan.

**2. Bab II Landasan Teori**

Bab ini memuat teori-teori dan penerapan metode yang relevan yang digunakan sebagai dasar atau acuan untuk memecahkan masalah yang ada dalam laporan tesis ini.

**3. Bab III Metodologi Penelitian**

Bab ini berisikan tentang langkah-langkah atau tahapan-tahapan yang akan dilakukan dalam melakukan proses penelitian.

**4. Bab IV Analisa dan Perancangan**

Pada bab ini menguraikan tentang analisa dan rancangan dalam melakukan pengolahan data dalam memprediksi jumlah pelanggan aktif

yang diperoleh saat penelitian, dan bagaimana penggunaan algoritma C4.5.

## **5. Bab V Implementasi dan Hasil**

Pada bab ini berisikan penjelasan mengenai penerapan dan penggunaan perangkat lunak yang telah dibuat.

## **6. Bab VI Kesimpulan dan Saran**

Berisikan kesimpulan dari hasil analisa serta perancangan sistem dan saran-saran untuk pengembangan di kemudian harinya.