

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pondok Pesantren Ngalah merupakan salah satu pondok pesantren di wilayah Kabupaten Pasuruan dengan jumlah 4000 santri berdasarkan data di kantor Yayasan Darut Taqwa. Dari jumlah santri yang sangat banyak tersebut banyak pula kebutuhan yang harus dikeluarkan oleh pesantren salah satunya adalah kebutuhan yang paling dasar seperti makanan, pesantren memberikan fasilitas berupa kos makan atau kupon makan yang bisa dibayarkan 1 bulan sekali, namun banyak juga santri yang terlambat melakukan pembayaran sampai berbulan-bulan, hal ini mengakibatkan kerugian pada pihak tertentu serta menurunnya pelayanan seperti menurunnya jumlah makanan yang disediakan karena kurangnya modal untuk berbelanja, akibatnya tidak sedikit santri yang memberikan keluhannya.

Dengan jumlah santri yang mencapai ribuan banyak pula data yang didapatkan. Data dalam jumlah besar dapat dikelola dan diproses dengan baik sehingga dapat menjadi sebuah kesempatan bagi instansi atau perusahaan, dengan adanya data dapat digunakan untuk memprediksi kejadian di masa yang akan datang (Kurniawan, 2018). Pengolahan data ini bisa dilakukan dengan proses data mining menggunakan metode klasifikasi dengan dua algoritma yang dibandingkan untuk menemukan tingkat akurasi yang paling tinggi.

Penelitian lain yang telah dilakukan dengan metode yang sama namun dengan satu algoritma yang dilakukan oleh (Istiana & Waspada, 2019) mengenai aplikasi pembayaran spp pondok pesantren dengan algoritma C4.5 yang menghasilkan rata-rata akurasi 81,15%, rata-rata nilai *precision* 77,62% dan rata-rata nilai *recall* 91,90%. Penelitian

lain yang dilakukan oleh (Muqorobin et al., 2019) mengenai prediksi keterlambatan pembayaran sumbangan pembinaan pendidikan sekolah dengan algoritma *Naive Bayes* menghasilkan akurasi 90%.

Kelebihan dari metode klasifikasi dengan algoritma C4.5 yang dilakukan Istiana, kinerja algoritma C4.5 lebih baik apabila dibandingkan dengan algoritma ID3, C5.0 dan algoritma CART. Penelitian kedua oleh Muqorobin dkk mengenai prediksi keterlambatan pembayaran sumbangan pembinaan pendidikan sekolah dengan algoritma *Naive Bayes* yang digabungkan dengan algoritma *Information Gain* menghasilkan akurasi lebih besar dibandingkan dengan menggunakan algoritma *Naive Bayes* tanpa penerapan algoritma *Information Gain* dengan selisih 10% keakurasiannya.

Menurut pendapat penyusun, metode yang tepat untuk mengukur tingkat kepuasan santri dengan cara klasifikasi, dengan membagi data kedalam variabel-variabel yang di tentukan yang kemudian diklasifikasikan menjadi atribut tepat atau terlambat membayar. Karena jumlah santri yang banyak maka banyak pula data yang dihasilkan sehingga perlu adanya penentuan algoritma yang tepat dalam memprediksi keterlambatan pembayaran kos makan. Algoritma yang akan di bandingkan dalam penelitian ini adalah algoritma C4.5 dan algoritma *Naive Bayes* yang tergolong dalam metode klasifikasi.

Penelitian ini dilakukan untuk menemukan algoritma dengan keakuratan tertinggi dari dua jenis algoritma dalam memprediksi keterlambatan pembayaran kos makan di pondok pesantren Ngalah. Penelitian ini dalam pelaksanaannya difokuskan di Asrama B dan Asrama C Pondok Pesantren Ngalah. Berdasarkan latar belakang masalah tersebut, maka penulis tertarik untuk melakukan

penelitian yang berjudul “PERBANDINGAN ALGORITMA *NAIVE BAYES* DAN C4.5 MENGGUNAKAN METODE KLASIFIKASI DALAM PREDIKSI KETERLAMBATAN PEMBAYARAN KOS MAKAN”.

1.2 Rumusan Masalah

1. Bagaimana cara memprediksi keterlambatan pembayaran kos makan di pondok pesantren Ngalah?
2. Bagaimana cara menentukan algoritma yang tepat dalam memprediksi keterlamabatan pembayaran kos makan di pondok pesantren Ngalah?

1.3 Tujuan

1. Untuk memprediksi keterlambatan pembayaran kos makan di pondok pesantren Ngalah.
2. Untuk menentukan algoritma yang tepat dalam memprediksi keterlamabatan pembayaran kos makan di pondok pesantren Ngalah.

1.4 Manfaat

1. Manfaat Teoritis
Berdasarkan tujuan di atas, maka manfaat yang diharapkan dalam penelitian ini adalah sebagai acuan untuk memprediksi keterlambatan pembayaran kos makan sehingga dapat mencari alternatif solusi untuk memecahkan masalah ini dengan melihat tingkat akurasi tertinggi dari dua algoritma yaitu C4.5 dan Naive Bayes.
2. Manfaat Praktisi
 - a. Mahasiswa
Mengetahui cara perhitungan C4.5 dan *Naive Bayes* dalam prediksi suatu permasalahan menggunakan data dengan jumlah besar.
 - b. Instansi

Menemukan akar permasalahan dari suatu prediksi, sehingga dapat mencari solusi permasalahan yang dialami.

1.5 Batasan Masalah

Penelitian ini dilakukan pada data keterlambatan pembayaran kos makan santri pondok pesantren ngalah yang difokuskan pada santri Asrama B dan Asrama C dengan jumlah 450 santri. Selain data keterlambatan pembayaran kos makan, penelitian ini difokuskan dalam membandingkan algoritma C4.5 dan Naive Bayes untuk penentuan tingkat akurasi didalam permasalahan ini.