

BAB I PENDAHULUAN

1.1 Latar Belakang

Customer Churn ialah keadaan dimana pelanggan berhenti berlangganan pada perusahaan tertentu. Seiring berkembangnya zaman, semua lebih mudah untuk mengakses segala hal. Potensi pelanggan untuk berpindah ke perusahaan lain menjadi tantangan bagi semua perusahaan agar bisa mengevaluasi strategi yang digunakan sebelumnya. Topik tentang *customer churn* ini memiliki daya tarik tersendiri untuk mengetahui pola dari pelanggan yang melakukan pemutusan kontrak atau berhenti berlangganan bagi suatu perusahaan. Di dalam lingkungan bisnis sangat diperlukan pemahaman mengenai mengapa pelanggan melakukan berhenti berlangganan (*Churn*) agar bisa bertahan dari berbagai perusahaan – perusahaan lain[1].

Bagi perusahaan telekomunikasi, secara khusus, sering menghadapi kehilangan pelanggan dan berpindah ke pesaing mereka. Dalam beberapa tahun terakhir perusahaan telekomunikasi sering kehilangan pelanggannya dikarenakan adanya pesaing baru, layanan yang lebih maju, dan teknologi yang lebih baru. Dengan kejadian tersebut tentu saja perusahaan mengalami kerugian yang cukup besar. Kejadian seperti ini juga dapat menimbulkan masalah yang serius bagi perusahaan tersebut[2].

Jika suatu perusahaan ingin bertahan dari perusahaan lain, namun mereka hanya mengandalkan *survey* kepuasan dari pelanggan itu saja belum cukup. Namun juga perlu diadakannya suatu cara baru untuk bahan evaluasi bagi suatu perusahaan[3]. Untuk menjaga loyalitas dari pelanggan menjadi tantangan tersendiri bagi suatu perusahaan. Pelanggan biasanya melakukan pemutusan kontrak dikarenakan suatu alasan, beberapa diantaranya seperti layanan suatu perusahaan tidak memenuhi ekspektasi pelanggan, kurangnya inovasi dari suatu perusahaan, dan juga bisa dikarenakan adanya penawaran yang lebih menarik perhatian pelanggan yang berasal dari kompetitor[4].

Semakin ketatnya persaingan di zaman sekarang, terutama di dunia industri, *customer churn* berhasil menjadi tantangan di berbagai sektor seperti perbankan, telekomunikasi, media, dan lain lain. *Customer Relationship Management* berperan sebagai pembangun hubungan dan mempertahankan hubungan jangka panjang

dengan pelanggan. Ini bertujuan untuk meningkatkan kepuasan pelanggan, pangsa pasar, dan juga keuntungan[5]. Hubungan ini juga diharapkan agar bisa di jalankan dalam jangka panjang. CRM ini diterapkan secara luas diberbagai sektor seperti telekomunikasi, perbankan, asuransi, dan lain – lain. Karena biaya untuk memperoleh pelanggan baru jauh lebih tinggi dibandingkan dengan mempertahankan pelanggan yang sudah ada. Maka dibutuhkannya program yang dapat memprediksi perilaku pelanggan yang *churn* untuk mengantisipasi peristiwa *customer churn*[6].

Dalam memprediksi pelanggan yang melakukan *churn*, tentu saja diperlukannya sistem yang efisien. Namun penggunaan pendekatan ini juga bisa berisiko tinggi ketika suatu program menghasilkan prediksi yang kurang akurat. Sehingga suatu perusahaan harus mengeluarkan biaya yang lumayan tinggi lagi untuk mencapai target yang mereka harapkan. Dalam dunia telekomunikasi *churn* ini merujuk kearah pemutusan kontrak dalam periode tertentu[7].

Agar perilaku *churn* ini bisa diketahui oleh suatu perusahaan, maka suatu perusahaan tersebut akan membuat suatu cara agar bisa memprediksi perilaku dari pelanggannya. Untuk mengevaluasi prediksi *churn* suatu pelanggan ini memakai *decision tree* sebagai metodenya, dan menggunakan algoritma C4.5. *Decision tree* ini sering dipakai untuk mengklasifikasi data *mining*. Cara kerjanya adalah dengan menggunakan pohon keputusan, dimana setiap jalur dari akar ke daun membentuk suatu aturan keputusan untuk menentukan kelas dari suatu baris data baru[8]. Setiap pohon memiliki beberapa cabang, cabang tersebut yang mewakili suatu kelas yang harus dipenuhi agar bisa menuju ke cabang selanjutnya[9].

Algoritma C4.5 ialah algoritma perkembangan dari algoritma ID3, peningkatan ini menggunakan *information gain* untuk memilih dari banyak atribut sebagai atribut pemisah. Pemilihan atribut pemisah ini adalah hal utama yang ada di dalam proses pembentukan pohon keputusan. Kriteria yang dipilih juga menentukan kinerja suatu struktur dan menentukan keputusan yang dihasilkan. Informasi gain ini diartikan sebagai entropi sebelum dan setelah perubahan[10].

Di dalam algoritma C4.5 setiap cabang akan diberi label dengan nilai dengan rentan tertentu. Untuk setiap simpul daun merupakan simpul akhir dari pohon yang memiliki label kelas. Disini *churn* sebagai atribut utama dan sebagai atribut

pembagi yang nantinya akan menghasilkan kelas “yes” dan “no”. Setiap jalur akan membentuk sebuah aturan yang terbentang dari akar ke simpul daun. C4.5 merupakan algoritma standar untuk menghasilkan suatu aturan klasifikasi dalam bentuk *decision tree*[11]. Dalam penelitian ini menggunakan dataset dari *Kaggle* yang berjudul “*Customer Churn Prediction*”.

1.2 Rumusan Masalah

Sesuai dengan penjabaran latar belakang yang sudah dijabarkan, dapat disimpulkan beberapa pernyataan yaitu.

1.2.1 Permasalahan

Dalam mengimplementasikan algoritma C4.5 *decision tree* ini menemukan beberapa permasalahan yang mempengaruhi kinerja sistem. Dikarenakan jumlah fitur yang digunakan terbilang banyak, yang membuat keputusan (cabang – cabang) dan hasil (daun – daun) dari pohon keputusan menjadi sangat banyak. Maka diperlukannya mempertimbangkan cara terbaik untuk membangun model *decision tree* agar nilai evaluasi menjadi akurat dalam klasifikasi *customer churn*.

1.2.2 Solusi Permasalahan

Solusi yang diusulkan untuk penelitian yaitu klasifikasi *customer churn* dengan metode algoritma C4.5 *decision tree* dengan melakukan pengujian berdasarkan fitur *churn*, dan menggunakan fitur *tenure*, *contract*, layanan tambahan, *payment method* sebagai patokan faktor yang paling mempengaruhi *churn*.

1.2.3 Pertanyaan Penelitian

Berapakah hasil algoritma C4.5 *Decision Tree* untuk mengklasifikasi *customer churn* dengan melakukan pengujian parameter data *testing* dan data *training*?

1.3 Batasan Masalah

Untuk memperoleh hasil yang diinginkan, jadi diperlukannya batasan masalah yang sesuai dengan topik yang telah dijabarkan sebelumnya. Berikut merupakan batasan – batasan masalah yang berhubungan dengan penelitian ini:

1. Penelitian ini menggunakan dataset yang diambil dari situs *Kaggle* (<https://www.kaggle.com/code/ybifoundation/telecom-customer-churn-prediction>).
2. Jumlah data yang digunakan berjumlah 7043 sampel data.
3. Jumlah fitur yang digunakan seluruhnya yaitu 21 fitur, diantaranya yaitu ID pelanggan, jenis kelamin, warga senior, status, tanggungan, masa berlangganan, layanan telepon, garis ganda, layanan internet, keamanan online, cadangan online, perlindungan perangkat, dukungan teknis, streaming TV, streaming film, kontrak, penagihan tanpa kertas, metode pembayaran, biaya bulanan, total biaya, dan berhenti langganan.
4. Metode yang digunakan *decision tree* dan menggunakan algoritma C4.5.
5. Klasifikasi yang digunakan berdasarkan *churn* atau *no churn*.

1.4 Tujuan dan Manfaat

Berdasarkan rumusan masalah yang telah diuraikan jadi tujuan dan manfaat dari penelitian ini yaitu.

1.4.1 Tujuan

Tujuan dari penelitian ini yaitu untuk mendapatkan hasil akurasi terbaik dengan menerapkan metode *decision tree* dan algoritma C4.5.

1.4.2 Manfaat

Manfaat dari penelitian ini yaitu untuk mengetahui akurasi yang dihasilkan dari penerapan metode *decision tree* untuk klasifikasi *customer churn* menggunakan dataset *telecom* yang ada di situs *Kaggle*.

1.5 Sistematika Laporan

Laporan ini disusun ke dalam lima bab dengan sistematis adalah dibawah ini:

BAB I PENDAHULUAN

Bagian ini berisi uraian tentang penelitian yang dilakukan. Ini juga mencakup latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, dan sistematika laporan.

BAB II TINJAUAN PUSTAKA

Bagian ini berisi landasan teori dan panduan yang menjadi dasar penelitian klasifikasi *customer churn* dengan metode algoritma *c4.5 decision tree*, dan mencakup studi literatur terkait dalam penelitian ini.

BAB III METODOLOGI PENELITIAN

Bagian ini berisi tentang arsitektur sistem, informasi dataset, preprosesing data, tahap – tahap dari metode yang diajukan, dan skenario pengujian.

BAB IV HASIL DAN PEMBAHASAN

Bagian ini berisikan uraian hasil dari penerapan sistem dan hasil dari pengujian skenario yang telah ditentukan.

BAB V PENUTUP

Bagian ini merupakan bagian terakhir didalam penelitian ini, berisikan simpulan dari penganalisisan yang telah diterapkan, serta memberi saran untuk pengembangan penelitian di masa depan.