

BAB 1

PENDAHULUAN

1.4. Latar Belakang

Klasifikasi adalah proses perencanaan objek ke dalam kelas, kelompok, atau kategori berdasarkan karakteristik yang telah ditentukan[1]. Klasifikasi pelanggan perlu dilakukan untuk mengetahui demografi pelanggan dengan menggunakan produk yang dapat dilihat dari berapa banyak pelanggan dan juga tingkat transaksi yang dilakukan[2]. Klasifikasi pelanggan secara efektif merupakan sebuah masalah utama yang harus diselesaikan perusahaan dalam hubungan manajemen dengan pelanggan sehingga mempengaruhi kepentingan perusahaan besar[3]. Perusahaan perlu memberikan kualitas pelayanan yang memuaskan agar pelanggan tidak beralih menggunakan merek lain dalam produk yang dipasarkan. Menggunakan teknik data mining merupakan metode yang dapat digunakan untuk menentukan kepuasan pelanggan[4].

Adanya masalah pada kecenderungan penurunan jumlah pelanggan speedy di wilayah operasional Sumatera Utara karena ketidakpuasan pelanggan. Pemutusan hubungan kerja dilakukan oleh pelanggan terhadap PT. Telekomunikasi Indonesia, Tbk di Sumatera Utara. Tidak ada pengelolaan klasifikasi data pelanggan sehingga informasi klasifikasi berdasarkan pembelian produk tertentu tidak dapat diketahui.

Teknik klasifikasi merupakan mekanisme analisis yang penting dalam memprediksi berbagai tingkat akurasi. Klasifikasi adalah salah satu metode dalam data mining untuk mengkategorikan kelompok barang tertentu ke dalam kelompok sasaran. Tujuannya adalah untuk memprediksi sifat suatu barang atau data berdasarkan kelas barang yang tersedia. Konstruksi model klasifikasi selalu ditentukan oleh kumpulan data pelatihan yang tersedia[5]. Data Mining menyediakan sekumpulan teknik untuk menemukan pola tersembunyi dari data[6].

Salah satu algoritma yang sering digunakan untuk mengklasifikasikan data menggunakan perhitungan probabilitas adalah Algoritma Naïve Bayes Classifier. Dari perhitungan tersebut, setiap kata dalam dokumen akan menghasilkan klasifikasi kategori

positif, negatif, dan netral yang diperoleh dari proses perhitungan sebelumnya yaitu pembobotan tf-idf[7]. Naïve Bayes merupakan algoritma klasifikasi yang mudah digunakan namun memiliki kelemahan yang mengakibatkan kinerja yang kurang baik, oleh karena itu diperlukan seleksi fitur, algoritma genetika merupakan algoritma yang mampu menyeleksi atribut dalam penelitian, akan dipilih berdasarkan bobot tertinggi sehingga agar akurasi hasil prediksi lebih optimal[8]. Algoritma Genetika adalah teknik untuk menemukan solusi optimal dari masalah yang memiliki banyak solusi. Teknik ini akan mencari dari beberapa solusi yang diperoleh untuk mendapatkan solusi terbaik sesuai dengan kriteria yang telah ditentukan atau yang dikenal dengan fungsi fitness. Algoritma ini termasuk dalam kelompok algoritma evolusioner yang menggunakan pendekatan evolusi Darwinian dalam bidang biologi seperti pewarisan, seleksi alam, mutasi gen dan kombinasi (crossover). Karena merupakan teknik pencarian yang optimal dalam bidang ilmu komputer, maka algoritma ini juga termasuk dalam kelompok algoritma metaheuristik.

1.4. Rumusan Masalah

Rumusan masalah dalam penelitian ini, antara lain:

1. Masalah yang sering muncul dalam pengelolaan data adalah tidak adanya pengelolaan klasifikasi data pelanggan menyebabkan informasi klasifikasi berdasarkan pada produk tertentu tidak dapat diketahui. Tidak adanya klasifikasi data juga dapat mempengaruhi manajemen dalam penentuan pengambilan keputusan.
2. Salah satu metode yang dapat digunakan adalah Naïve Bayes Classifier dengan pemilihan fitur Algoritma Genetika untuk melakukan klasifikasi data pelanggan, sehingga dalam sebuah proses pengolahan data, perusahaan dapat mengklasifikasi tiap data pelanggan secara akurat.

1.3. Tujuan dan Manfaat

1.3.1. Tujuan

Tujuan penelitian ini adalah untuk memprediksi potensial pelanggan dan non pelanggan dengan mengklasifikasi terhadap data pelanggan pada PT. Telekomunikasi Indonesia, Tbk dalam memasarkan produk barunya yaitu Indihome.

1.3.2. Manfaat

Manfaat yang diharapkan dari penelitian ini antara lain:

1. Menghasilkan laporan mengenai klasifikasi menggunakan Naïve Bayes Classifier dengan pendekatan Algoritma Genetika
2. Menghasilkan laporan mengenai performa Naïve Bayes Classifier dalam memprediksi klasifikasi kelompok data pelanggan.
3. Dengan penelitian Naïve Bayes Classifier yang dilakukan oleh penulis, diharapkan dapat memberi kontribusi yang berguna dalam perkembangan teknologi pada masa yang akan datang dalam melakukan proses klasifikasi data yang dimiliki oleh perusahaan, maupun pemerintahan.

1.4. Batasan Masalah

Ruang lingkup penelitian ini, adalah sebagai berikut:

1. Jumlah data pelanggan Speedy dan Indihome yang diteliti sebanyak 878 data pelanggan, diantaranya memakai sebanyak 200 data sebagai data klasifikasi pelanggan.
2. Algoritma machine learning yang digunakan adalah Naïve Bayes Classifier, dengan pemilihan fitur Algoritma Genetika (GA-NBC).
3. Perangkat lunak yang digunakan adalah software Rapidminer 5.3.

1.5. Keterbaruan

Model Naive Bayes adalah sekelompok algoritma klasifikasi yang sangat cepat dan sederhana yang seringkali cocok untuk kumpulan data berdimensi sangat tinggi. Karena mereka sangat cepat dan hanya memiliki sedikit parameter yang dapat disetel, mereka akhirnya menjadi sangat berguna sebagai dasar cepat dan kotor untuk masalah

klasifikasi[9]. Hasil klasifikasi komparatif menggunakan algoritma naïve bayes dan C4.5 menyatakan bahwa algoritma naïve bayes merupakan model klasifikasi terbaik yang dapat digunakan untuk klasifikasi pelanggan IndiHome untuk menentukan pelanggan potensial dan non potensial [10]. Algoritma genetika untuk mengoptimalkan pemilihan fitur pada algoritma Naive Bayes agar akurasi meningkat. Mampu meningkatkan akurasi pada Naive Bayes dengan mengoptimalkan proses pemilihan fitur berupa kata-kata pada abstrak setiap penelitian. Hasil eksperimen pada penelitian ini menunjukkan bahwa nilai akurasi meningkat sebesar 26,06% dengan menggunakan Algoritma Genetika pada proses seleksi fitur[11]. Penelitian sebelumnya menggunakan Naïve Bayes memiliki tingkat akurasi hanya sampai 69,60% setelah dilakukan seleksi fitur dengan algoritma genetika tingkat akurasi meningkat menjadi 96,67%, selisih peningkatan akurasi sebesar 27,07% [8].

Pola pembelajaran algoritma genetika menginformasikan pendatang baru atau klasifikasi baru dengan waktu yang lebih cepat. Kesulitan konfigurasi di naïve Bayes dapat dibantu oleh algoritma genetika dan menyediakan pemodelan yang memadai untuk mendeskripsikan sistem. Penelitian tersebut menghasilkan nilai akurasi sebesar 81,19%, terbukti bahwa algoritma Naïve Bayes dapat meningkatkan kinerjanya dengan algoritma genetika pada klasifikasi bakteri [12]. Implementasi metode naïve bayes dan seleksi fitur berbasis algoritma genetika untuk klasifikasi dokumen SAMBAT online menghasilkan rata-rata akurasi sebesar 86,12% dengan nilai akurasi tertinggi sebesar 89,79% dalam 5 kali percobaan dengan menggunakan 49 data uji dan dengan nilai parameter terbaik yaitu banyak generasi 70, ukuran populasi 20, tingkat persilangan 0,8 dan tingkat mutasi 0,2 [13]