

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada tahun 2019 Asosiasi Jasa Internet Indonesia (APJII) menyebut bahwa pengguna internet di Indonesia mencapai 196,7 juta atau setara dengan 73,7 persen dari jumlah total penduduk Indonesia mencapai 266.911.900 menurut data proyeksi Badan Pusat Statistik (BPS) pada tahun 2019. Angka tersebut naik dari 64,8 persen dibanding tahun 2018 (Slamet JP, 2021). Oleh karenanya berbagai perusahaan di bidang telekomunikasi bersaing menyediakan pelayanan terbaik dan menciptakan berbagai inovasi guna memenuhi kebutuhan konsumen yang semakin kompleks seiring perkembangan jaman.

Salah satu perusahaan di bidang telekomunikasi adalah PT Telkom Indonesia. PT Telkom (Persero) Tbk (Telkom) merupakan Badan Usaha Milik Negara (BUMN) yang bergerak di bidang jasa layanan teknologi informasi dan komunikasi serta jaringan telekomunikasi di Indonesia. Indihome (*Indonesia Digital Home*) adalah salah satu layanan dari PT Telkom yang berupa paket layanan telekomunikasi data yaitu telepon rumah (*voice*), internet (*internet on fiber*), dan layanan tv (*use tv cable*) (Lestari & Uly Artha, 2017). Disisi lain PT Telkom juga terus memperluas jaringan fiber optik demi melayani kebutuhan internet para pelanggannya. Namun dari itu semua tetap ada konsekuensinya yaitu apabila layanan Indihome mengalami gangguan maka kepuasan pelanggan akan berkurang.

Kepuasan pelanggan merupakan respon dari pemenuhan kebutuhan pelanggan. Kepuasan pelanggan juga didefinisikan sebagai wujud perasaan pelanggan setelah membandingkan dengan harapannya. Apabila perusahaan mempunyai kinerja di bawah harapan konsumen maka konsumen akan kecewa begitu pula sebaliknya (Ong & Pambudi, 2014).

Selama ini PT. Telkom mengetahui review pelanggannya melalui telepon customer service atau melalui komplain langsung, sedangkan PT. Telkom belum mengetahui review pelanggan dari media social seperti twitter. Penelitian ini difokuskan untuk melakukan analisis terhadap isi *tweet* pengguna twitter tentang pengalaman mereka selama menggunakan layanan

indihome. Metodologi pengumpulan data yang digunakan adalah dengan salah satu teknik text mining yang biasa disebut dengan teknik *Crawling* data dengan kata kunci “Indihome”.

Twitter adalah sebuah sosial media yang memiliki fitur *micro-blogging* (blog berukuran kecil) dengan dengan jumlah karakter dalam tweet (post dalam twitter) maksimal 140 karakter. Pada tahun 2016 jumlah pengguna mencapai 24,3 juta orang (Rizkia et al., 2019). Twitter seringkali dijadikan tempat untuk menyampaikan opini seperti terhadap tokoh publik, layanan publik, maupun produk dari sebuah perusahaan, tidak terkecuali terhadap produk indihome. Banyak pengguna twitter yang juga merupakan pelanggan indihome menyampaikan opini mereka terhadap produk indihome baik berupa opini yang bersifat positif, netral, maupun negatif (Rizkia et al., 2019). Teknik yang biasa digunakan dalam mengunduh isi tweet ataupun atribut lain yang ada di twitter adalah teknik *Crawling* Data.

Crawling data di Twitter merupakan suatu proses untuk mengunduh atau mengambil data dari server twitter menggunakan bantuan *Application Programming Integration (API)* twitter baik berupa data *user* maupun data tweet. Beberapa penelitian sebelumnya menggunakan data yang didapat dari twitter sebagai bahan acuan untuk penelitian mereka (Eka Sembodo et al., 2016).

Klasifikasi merupakan bagian dari prediksi, yang mana nilainya berupa label. Klasifikasi akan menentukan *class* atau grup untuk masing-masing contoh data. Input dari klasifikasi adalah atribut/variabel yang berasal dari data sample, sedangkan output dari klasifikasi adalah *class* dari data sample itu sendiri. Dalam machine learning, klasifikasi digunakan untuk metode supervised learning. Metode supervised learning adalah cara yang digunakan untuk menemukan hubungan antara atribut masukan dengan atribut target, selanjutnya akan ditemukan sebuah hubungan yang disebut sebagai model (Frastian et al., 2018). Dalam klasifikasi kita dapat menemukan orang atau objek kedalam suatu kategori tertentu, contoh untuk masalah klasifikasi yaitu menemukan apakah pelanggan “puas” atau “tidak puas” terhadap suatu produk tertentu. Informasi tentang pelanggan sebelumnya digunakan sebagai bahan untuk melatih algoritma dengan tujuan untuk mendapatkan rule atau aturan.

Seperti pada penelitian yang dilakukan oleh Shakina Rizkia dkk (Rizkia et al., 2019), dalam penelitian tersebut dilakukan pengambilan data dengan teknik *Crawling* data dari twitter untuk melakukan analisis sentimen kepuasan pelanggan terhadap layanan Indihome

menggunakan metode *decision tree ID3* yang menghasilkan akurasi maksimal sebesar 80,1% dengan pembobotan menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)*.

Erlin Elisa (Elisa, 2017) juga melakukan penelitian dengan menerapkan metode Algoritma C4.5 untuk mengidentifikasi faktor-faktor penyebab kecelakaan kerja. Data didapat dari data inspeksi K3 (Keselamatan dan Kesehatan Kerja Konstruksi) kemudian data diolah menggunakan proses *Knowledge discovery in database (KDD)*. Penelitian tersebut menghasilkan kesimpulan bahwa faktor kecelakaan kerja disebabkan oleh pekerja dan cara kerja, lingkungan tempat kerja, dan alat pelindung diri.

Maka dalam penelitian ini, penulis mengusulkan metode *Decision tree* Algoritma C4.5. Algoritma C4.5 merupakan salah satu metode klasifikasi. Algoritma C4.5 adalah algoritma yang digunakan untuk membuat pohon keputusan. Pohon keputusan dikenal sebagai metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Dengan menggunakan metode pohon keputusan fakta yang sangat besar dapat diubah menjadi suatu keputusan yang merepresentasikan aturan. Sehingga aturan dapat mudah dipahami dengan Bahasa alami (Mochamad Rizki Ilham, 2016).

Alasan penggunaan metode algoritma C4.5 dikarenakan metode ini dikenal sebagai metode yang memiliki tingkat akurasi yang tinggi. Seperti pada penelitian yang dilakukan oleh Achmad Rifai dan Rizki Aulianita (Rifai & Aulianita, 2018), yang mengkomparasikan metode Algoritma C4.5 dengan metode *Naïve Bayes*. Pada penelitian tersebut nilai akurasi tertinggi didapatkan oleh metode Algoritma C4.5 yaitu sebesar 87,61% dan disimpulkan bahwa metode algoritma C4.5 adalah metode terbaik dan dapat digunakan untuk pemecahan masalah kelayakan penentuan kredit pada koperasi. Selanjutnya pada penelitian yang dilakukan oleh Nahot Frastian dkk (Frastian et al., 2018). Pada penelitian tersebut bertujuan untuk mengkomparasikan metode algoritma C4.5 dengan metode *Naïve Bayes* dan metode *Random Forest* menghasilkan kesimpulan bahwa Algoritma C4.5 mempunyai akurasi tertinggi yaitu sebesar 98,89%.

Kedepan harapan dari peneliti adalah mampu mengidentifikasi faktor-faktor yang mempengaruhi kepuasan pelanggan PT Telkom sehingga hasil dari penelitian ini dapat menjadi pertimbangan bahkan menjadi acuan bagi para pemangku kebijakan di PT Telkom untuk senantiasa mewujudkan kepuasan pelanggannya.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Sebagai perusahaan telekomunikasi yang sangat populer di Indonesia, PT. Telkom perlu mewujudkan kepuasan pelanggan dengan tujuan terciptanya loyalitas pelanggan. Maka dari itu diperlukan analisis sentimen kepuasan pelanggan untuk mengidentifikasi faktor-faktor kepuasan pelanggan terhadap PT. Telkom khususnya layanan indihome. Permasalahan yang timbul adalah bagaimana menganalisis faktor-faktor kepuasan pelanggan PT. Telkom khususnya terhadap layanan indihome.

1.2.2 Metode Usulan

Penelitian ini mengusulkan metode Algoritma C4.5 sebagai metode pengklasifikasian untuk analisis sentimen kepuasan pelanggan, metode KDD (*Knowledge discovery in database*) sebagai metode untuk mengolah data. Sedangkan pengambilan data dilakukan menggunakan metode *Crawling* data twitter. Data yang sudah diklasifikasikan kemudian akan dilakukan pengujian menggunakan metode *Confusion Matrix*.

1.2.3 Pertanyaan Penelitian

- 1) Berapa akurasi C4.5 untuk analisis sentimen kepuasan pelanggan PT Telkom?
- 2) Bagaimana perangkingan variabel kepuasan pelanggan PT. Telkom?
- 3) Berapa presentase tweet tentang indihome yang bersifat puas, netral, dan tidak puas?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Berdasarkan rumusan masalah diatas maka penelitian ini bertujuan sebagai berikut:

- 1) Mengetahui akurasi algoritma C4.5 dalam analisis sentimen kepuasan pelanggan PT. Telkom.
- 2) Mengetahui perangkingan variabel mana yang paling mempengaruhi kepuasan pelanggan PT.Telkom.
- 3) Mengetahui presentase tweet yang termasuk dalam kelas puas, netral, dan tidak puas.

1.3.2 Manfaat Penelitian

Setelah penelitian ini selesai dilakukan dan faktor-faktor kepuasan pelanggan dapat diidentifikasi, penulis berharap penelitian ini akan menghasilkan manfaat yaitu sebagai referensi untuk penelitian-penelitian selanjutnya yang berhubungan dengan topik analisis kepuasan pelanggan dan dapat dijadikan acuan pendukung bagi para pemangku kebijakan di PT. Telkom untuk memperbaiki layanannya.

1.4 Batasan Masalah

Agar pembahasan lebih terarah, maka batasan masalah dalam penelitian ini adalah sebagai berikut.

- 1) *Dataset* yang digunakan dalam penelitian ini adalah 1000 dokumen *tweet* berbahasa indonesia yang di-*tweet* selama periode 1 Januari 2021 hingga 30 Juni 2021, dengan kata *indihome* sebagai topik.
- 2) *Crawling* dilakukan menggunakan bahasa pemrograman *Python* dan *Software Jupyter notebook*.
- 3) Pelabelan *tweet* berdasarkan tiga kategori tingkatan kepuasan yaitu puas, netral, dan tidak puas dilakukan secara manual oleh penulis kemudian divalidasi oleh dosen pembimbing. Sedangkan faktor-faktor kepuasan yang dijadikan sebagai atribut meliputi produk, pelayanan, harga, dan jaringan internet.

1.5 Sistematika Penulisan Proposal

Adapun sistematika penulisan proposal skripsi dari penelitian ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Bab ini merupakan bagian awal yang berisi latar belakang masalah mengapa identifikasi faktor-faktor kepuasan pelanggan PT Telkom menarik untuk diteliti, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan proposal skripsi.

BAB II. KAJIAN PUSTAKA

Bab ini memuat tinjauan pustaka yang menjadi dasar penelitian. Teori yang dimuat merupakan teori yang berpijak pada bidang *Data Mining* dengan menggunakan metode

algoritma C4.5. Serta beberapa penelitian sebelumnya yang dijadikan acuan dalam melakukan penelitian ini.

BAB III. PERANCANGAN SISTEM

Bab ini berisikan deskripsi tentang bagaimana penelitian akan dilaksanakan dengan menggunakan metode algoritma C4.5, jenis dan sumber data, metode klasifikasi data, serta metode analisis.

REFERENSI

Memuat daftar-daftar referensi atau sumber jurnal-jurnal yang digunakan sebagai acuan pada penelitian ini.

1.6 Perkiraan Jadwal Penelitian

Penelitian dijadwalkan membutuhkan waktu enam bulan. Adapun rincian tahapan penelitian adalah dijelaskan dalam tabel sebagai berikut.

Tabel 1. 1 Jadwal penelitian

No.	Kegiatan	Bulan 1				Bulan 2				Bulan 3				Bulan 4				Bulan 5				Bulan 6			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.	Mempelajari penelitian tentang analisis kepuasan pelanggan																								
2.	Mencari referensi terkait penelitian kepuasan pelanggan																								
3.	Mencari referensi																								

