

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Di era *society* 5.0 seperti sekarang ini, seringkali masyarakat mencari informasi dengan media sosial sebagai akses utama, seperti aplikasi twitter yang sering digunakan di kalangan masyarakat. Tingkat penggunaan internet di kalangan masyarakat sangat tinggi terutama dalam penggunaan media sosial. Media sosial yang sering dioperasikan dengan tujuan mencari semua informasi salah satunya ialah twitter[1]. Di media sosial twitter sering digunakan sebagai wadah untuk menyampaikan opini masyarakat. Maka dari itu, seringkali perusahaan menggunakan media sosial sebagai jejaring sosial untuk melakukan pemasaran bisnis salah satunya perusahaan yang bergerak di bidang layanan provider indihome.

Analisis sentimen merupakan proses untuk mengetahui suatu pendapat atau opini yang disampaikan secara tertulis baik berupa teks maupun dokumen apakah mempunyai sentimen positif atau negatif[2]. Penelitian yang berhubungan dengan analisis sentimen belakangan ini menjadi perhatian karena menghasilkan data seperti opini masyarakat, promosi politik, dan pengambilan keputusan [3]. Tujuan dari analisis sentimen adalah menganalisis suatu komentar dan mengelompokkannya sebagai suatu sentimen positif atau negatif[4]. Analisis sentimen itu sendiri dapat di terapkan dalam pengklasifikasian opini-opini atau *tweets* masyarakat di twitter terhadap kualitas layanan *provider* indihome di indonesia.

Indihome saat ini mengalami permasalahan dalam jaringan sehingga banyak keluhan dari pelanggan. Oleh karena itu, dalam peningkatan kualitas pelayanan dan jaringan perlu ditingkatkan lagi. Jika pelayanan indihome lebih ditingkatkan lagi kualitasnya maka pelanggan akan merasa puas, bahkan pelanggan tidak mudah beralih dan tidak menggunakan layanan yang lain. Tidak sedikit perusahaan operator yang menyediakan layanan *broadband* dan mereka berkompetisi dalam menyuguhkan harga serendah-rendahnya, namun luput dari pertimbangan akan kualitas layanan yang seyogyanya tonggak utama bagi perusahaan jasa. Salah satu layanan Indihome yang sering mendapatkan respon positif maupun negatif adalah layanan kualitas jaringan bahkan cakupan wilayah. Melalui media sosial twitter, pelanggan dapat mengutarakan atau mengkritik kualitas dari pelayanan indihome.

Berdasarkan banyaknya pendapat yang ada di twitter, jika diteliti lebih detail lagi akan menghasilkan analisis sentimen yang bermanfaat dalam mengetahui seberapa puas

konsumen pada produk serta layanan yang ada. Dengan adanya menganalisis pendapat pelanggan bisa dijadikan sebagai tahap proyeksi bagi layanan indihome agar lebih ditingkatkan lagi kualitas layanannya. Ada 2 tujuan dari menganalisis sebuah komentar yaitu mengelompokkan komentar positif dan negatif. Terdapat beberapa model algoritma pengklasifikasian data dalam melaksanakan analisis sentimen untuk diterapkan, yakni algoritma *Naïve Bayes*, dan *K-Nearest Neighbors (KNN)*, *Decision Tree C4.5*[2].

Model algoritma *Naive Bayes Classifier* diterapkan sebagai model pada penelitian saat ini, yang tujuannya mengklasifikasikan analisis sentimen pada aplikasi twitter tersebut. Tergolong sebagai algoritma yang cukup sederhana dengan mengkalkulasikan peluang, bisa juga dengan data yang muncul dikalkulasikan serta satu set dengan nilai-nilai gabungan[2]. Data yang berupa teks di twitter akan diseleksi terlebih dahulu lalu dihitung permasalahannya menggunakan *naïve bayes*. Sebuah data dari twitter yang berupa teks akan di ekstrak dan dibobotkan sebelum melakukan perhitungan algoritma *naïve bayes*. Tf-idf merupakan Ekstraksi fitur yang digunakan dalam *naïve bayes* untuk pembobotan kata. Tf-idf yakni suatu metode ekstraksi fitur dan pembobotan sebuah kata guna mengetahui suatu nilai kata dalam sebuah dokumen atau kalimat yang berupa data[5]. Dalam melakukan ekstraksi fitur dari tf-idf akan menghasilkan sebuah data yang sangat banyak, dengan demikian dibutuhkannya seleksi fitur guna menentukan atribut-atribut atau beberapa fitur yang paling penting.

Mengkaji dari penelitian terdahulu, dengan salah satu judul *Analisis Sentimen Kepuasan Pelanggan Terhadap Internet Provider Indihome di Twitter Menggunakan Metode Decision Tree Dan Pembobotan TF-IDF*[2]. Bagi penulis metode yang digunakan tersebut telah menghilangkan informasi ketika berupaya mengkategorikan variabel kedalam kategori-kategori. Mengetahui dari tujuannya dari metode *decision tree* yakni upaya memprediksi hasil dari variabel yang sifatnya *kontinu*, dan metode tersebut dianggap tidak efektif. Maka dari itu, penulis menerapkan metode *Naïve Bayes* karena penggolongannya berlandaskan probabilitas yang tidak rumit serta disusun untuk diimplementasikan sesuai sifatnya yang independensi. Adapun layanan *provider* indohome sebagai objek penelitian saat ini, merupakan salah satu akses sentral dalam beraktivitas sehari-hari, mulai dari pekerjaan, proses pembelajaran dsb. Seperti yang kita rasakan bersama adanya pandemi covid-19 menjadikan sebuah problematika serius yang harus kita hadapi. Perlu diketahui

bersama juga, jika adanya layanan indihome tersebut justru meringankan beban orang dalam menjalankan realitas kehidupannya, seperti meminimalisir *budget* jika dibandingkan dengan pembelian paket data harian, mingguan, bahkan bulanan.

Untuk pengumpulan data dari twitter dengan *crawling* data yang dilanjutkan dengan perhitungan algoritma *naïve bayes* dengan menentukan kelasnya dan labelnya. Untuk proses melakukan sebuah *crawling* data dengan menentukan sebuah kelas dari suatu kalimat dan sebuah akurasi dari algoritma *naïve bayes*. Untuk penelitian sebelumnya, tingkat *naïve bayes* mencapai 75,42%[6]. Dalam penelitian analisis sentimen ini diharapkan akan memberikan informasi yang bisa dijadikan acuan untuk lebih ditingkatkan lagi kualitas layanannya terutama perusahaan yang bergerak di bidang layanan *provider* indihome.

## **1.2 Perumusan Masalah**

### **1.2.1 Permasalahan**

Kualitas jaringan dan layanan *provider* indihome mengundang responsif atau sentimen pada masyarakat, salah satunya pada aplikasi twitter. Indihome saat ini mengalami permasalahan dalam jaringan dan layanan sehingga banyak keluhan dari pelanggan. Adapun responsifitas yang diterima bermacam-macam bentuknya, baik itu dari sisi tanggapan negatif hingga tanggapan positif. Oleh karena itu, dibutuhkannya analisis sentimen guna mendapatkan informasi yang dimaksud, lebih mendominasi pada sentimen negatif atau positif dari kalangan masyarakat. Pengklasifikasian sentimen yang berupa data teks dapat menggunakan perhitungan algoritma *naïve bayes*. Namun, terlebih dahulu mengekstraksi data yang bentuknya teks, kemudian *td-idf* dijadikan sebagai metode untuk melakukan pembobotan kata. Fitur yang dihasilkan dari proses *tf-idf* dan diseleksi untuk memilih fitur-fitur yang relevan menggunakan metode *information gain*.

### **1.2.2 Metode Usulan**

Peneliti merekomendasikan sekaligus menerapkan metode *naive bayes classifier* sebagai algoritma pengkategorian sentimen, dan untuk mengekstraksi fitur dan pembobotan katanya digunakan metode *tf-idf*, metode *information gain* digunakan sebagai seleksi fitur, dan untuk evaluasi model digunakan metode *k-fold cross validation*.

### **1.2.3 Pertanyaan Penelitian**

1. Berapa nilai akurasi model terbaik yang dihasilkan sistem algoritma *naïve bayes* yang diterapkan untuk mengklasifikasikan data tentang kualitas layanan *provider*

indihome?

2. Bagaimana analisis sentimen kualitas layanan *provider* indihome dari data twitter menggunakan model terbaik?

### **1.3 Tujuan dan Manfaatnya**

#### **1.3.4 Tujuan**

1. Mengetahui akurasi pada model algoritma sistem *naïve bayes* untuk mengklasifikasikan data yang dihasilkan
2. Untuk mengetahui hasil sentimen publik pada komentar twitter terkait kualitas layanan *provider* indihome di Indonesia digunakan Algoritma *Naive Bayes Classifier* (NBC).

#### **1.3.5 Manfaat**

1. Mendapatkan informasi terkait bagaimana opini masyarakat pengguna twitter terhadap kualitas layanan *provider* indihome.
2. Untuk dijadikan sebagai media penunjang dalam memperoleh masukan dari masyarakat dalam upaya membuat keputusan.

### **1.4 Batasan Masalah**

Penetapan batasan masalah dilaksanakan guna tidak terjadi pelebaran maupun penyimpangan pembahasan pokok masalah penelitian sehingga penelitian menjadi lebih fokus, mudah dalam menguraikan pembahasan, dan penelitian terwujudkan atas dasar tujuannya. Adapun beberapa batasan masalah pada penelitian kali ini yakni diantaranya :

1. Seputar penelitian hanya dalam analisis sentimen *user* twitter terhadap kualitas layanan *provider* indihome dengan kata kunci “*kualitas layanan indihome*”.
2. Data yang diperlukan adalah data dari twitter sebanyak 1136 data.
3. Perhitungan TF-IDF, *Information gain*, dan *Naïve Bayes*.

### **1.5 Metodologi Penulisan Skripsi**

Adapun sistematika penulisan pada penelitian ini merujuk pada pedoman karya tulis ilmiah yang ditentukan oleh Universitas Trunojoyo Madura:

#### **BAB I. PENDAHULUAN**

Di bagian bab ini menguraikan perihal cikal bakal munculnya , rumusan

masalah, kemudian manfaat dan tujuan, beberapa batasan masalah, dan juga alur ketentuan pembahasan.

## BAB II. TINJAUAN PUSTAKA

Mendeskripsikan terkait tinjauan pustaka serta rancangan teori yang berelasi terhadap tema yang ditentukan.

## BAB III. PERANCANGAN SISTEM

Pada bab membahas tentang perancangan sistem yang akan dibuat meliputi tahap studi pustaka, analisa dan perancangan sistem, dan implementasi sistem.

## BAB IV. HASIL PENELITIAN DAN ANALISA

Pada bab ini dijelaskan mengenai implementasi sistem yang dikerjakan sesuai dengan rancangan yang dibuat pada bab sebelumnya.

## BAB V. PENUTUP

Di bagian bab terakhir tersebut, pemaparan terkait kesimpulan dari hasil penelitian, baik dari pengaplikasian, pengujian yang berulang, dan juga tertulis saran atas pelaksanaan penelitian berupa skripsi ini oleh penulis.

Sistem yang dibangun dalam penelitian ini masih memiliki kekurangan yaitu data dalam *stopword* dan *slangword* yang digunakan dalam tahapan preprocessing data yang masih belum sempurna dan akan berpengaruh terhadap perhitungan algoritma *naïve bayes* sekaligus tingkat akurasi model. Penelitian selanjutnya diharapkan dapat lebih memperhatikan mengenai data *stopword* dan *slangword* sehingga model yang dibangun akan lebih baik.