

BAB I

PENDAHULUAN

1.1 Latar Belakang

Social media merupakan sebuah media berbasis internet yang digunakan oleh sebagian besar individu dalam melakukan interaksi terhadap sesama dengan cara sharing atau bertukar ide maupun informasi dalam suatu jaringan. Saat ini, *social media* sudah banyak digunakan oleh kalangan remaja di berbagai belahan dunia baik dari anak-anak, usia remaja hingga orang dewasa. Total pengguna *social media* di Indonesia [1] sendiri telah mencapai 150 juta pengguna yang berarti sudah mencapai 56% dari jumlah total penduduk di Indonesia.

Salah satu *social media* yang banyak digunakan oleh masyarakat luas adalah *Twitter*. Pada kuartal ke-3 tahun 2019 [2], pengguna aktif harian *Twitter* telah meningkat sebesar 17 persen hingga mencapai 145 juta pengguna. *Twitter* merupakan platform *social media* dimana penggunanya dapat dengan mudah mengemukakan opini yang beberapa diantaranya terdapat informasi yang sangat berharga mengenai tokoh publik, layanan publik maupun produk dari sebuah perusahaan.

Seiring dengan maraknya penggunaan *Twitter* tersebut, *social media* marketing melalui *Twitter* menjadi salah satu kunci strategis dalam kegiatan marketing suatu brand di Indonesia bahkan di dunia. Salah satu perusahaan yang memanfaatkan *Twitter* sebagai platform untuk marketing yaitu Indihome, yang merupakan perusahaan penyedia layanan internet di Indonesia. Saat ini, Indihome

telah memiliki banyak pelanggan sehingga dengan besarnya angka pengguna layanan tersebut, maka dapat memunculkan berbagai opini khususnya pada platform twitter, baik yang bersifat pujian maupun keluhan. Dengan munculnya berbagai opini tersebut, apabila dikumpulkan dan diolah dapat dilakukan analisis sentimen untuk mengetahui apakah pelanggan puas atau tidak terhadap penyedia layanan internet Indihome. Analisis sentimen sangat diperlukan guna menyaring komentar yang kurang baik pada sebuah sosial media.

Analisis sentimen merupakan [3] sebuah studi komputasional yang berasal dari beberapa sentimen, opini dan emosi orang dengan melalui entitas dan atribut yang dimiliki dan diekspresikan menjadi bentuk teks. Sentimen analisis merupakan bagian dari *Natural Language Processing* dan *Machine Learning* [4] cara kerjanya yaitu dengan mengklasifikasi opini positif, negatif, dan netral. Pada umumnya, penelitian mengenai analisis sentimen sudah sangat sering dilakukan, salah satunya yaitu [5] analisis sentimen yang mengklasifikasi review film dengan menggunakan *machine learning* dan menghasilkan klasifikasi positif dan negatif.

Analisis sentimen memerlukan sebuah metode yang dapat menunjang proses klasifikasi. Salah satu metode klasifikasi yang dapat digunakan yaitu *K-Nearest Neighbor*. Metode KNN merupakan [6] salah satu metode yang digunakan dalam proses klasifikasi secara mudah dan efisien. Algoritma *K-Nearest Neighbor* (KNN) [7] merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Contoh penelitian yang menggunakan metode KNN adalah prediksi harga jual tanah yang dilakukan oleh Wiyli Yustanti. Peneliti menggunakan [7] metode

gabungan antara tahapan datang mining *CRISP-DM* dan metode pengembangan perangkat lunak *Waterfall Model*. Kemudian dari penelitiannya tersebut didapatkan hasil akurasi dari prediksi data testing sebesar 80%.

Berdasarkan latar belakang yang telah diuraikan di atas, penulis akan meneliti mengenai kepuasan pelanggan pada akun twitter Indihome dengan menggunakan algoritma *K-Nearest Neighbor*. Hasil penelitian tersebut akan dibagi menjadi dua kelas sentiment, yaitu positif dan negatif.

1.2 Rumusan Masalah

Adapun yang menjadi rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimanakah proses klasifikasi sentimen kepuasan pelanggan pada akun twitter Indihome dengan menggunakan algoritma *K-Nearest Neighbor*?
2. Berapakah nilai akurasi yang dihasilkan dari metode *K-Nearest Neighbor* pada sentimen analisis yang dilakukan?

1.3 Batasan Masalah

Berdasarkan rumusan masalah di atas, maka batasan masalah dalam penelitian ini adalah:

1. Penerapan algoritma *K-Nearest Neighbor* dalam menganalisis sentiment kepuasan pelanggan pada akun twitter Indihome.
2. Data yang akan digunakan adalah kata berbahasa Indonesia yang berhubungan dengan Indihome.
3. Analisis sentiment akan diklasifikasikan menjadi dua kelas yaitu positif dan negatif.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menerapkan algoritma *K-Nearest Neighbor* dalam mengklasifikasikan sentimen sekaligus mengetahui seberapa besar nilai akurasi sentiment kepuasan pelanggan pada akun twitter Indihome

1.5 Manfaat Penelitian

Adapun manfaat pada penelitian ini adalah membantu Indihome untuk mengetahui sentiment apakah yang berikan oleh pelanggan terhadap pelayanan yang telah dilakukan, sehingga dapat digunakan sebagai tolak ukur dalam melakukan evaluasi pada proses bisnis dan kegiatan operasional perusahaan.

1.6 Metode Penelitian

1.6.1 Metode Pengumpulan Data

Metode ini akan dilakukan dua tahap yaitu studi pustaka dan *crawling* data.

1.6.1.1 Studi Pustaka

Pada metode ini dilakukan pencarian materi dan data yang bersumber dari buku, internet, jurnal, maupun referensi lainnya yang sesuai dan bersifat valid atau terpercaya sehingga dapat membantu dalam pelaksanaan dan penyusunan skripsi.

1.6.1.2 Crawling Data

Metode ini merupakan metode untuk mengambil data dari twitter sesuai dengan kata kunci apa yang akan dicari. Proses *crawling* data dari twitter dapat dilakukan menggunakan API Twitter, Bahasa pemrograman python, twitterscraper, atau dengan web.

1.6.2 Metode Analists

Pada metode ini dilakukan beberapa tahapan seperti:

1. *Data Collecting*

Data twitter akan *dicrawling* menggunakan software Jupyter Lab, dengan tahap sebagai berikut:

- a. Pertama yang dilakukan adalah meminta izin twitter dan mendapatkan *keys* dan *tokens* pada API twitter
- b. Kemudian menginstal *library* yang dibutuhkan
- c. Menginputkan kode yang didapatkan kemudian running program
- d. Selanjutnya menginputkan kata apa dan berapa banyak yang akan diambil
- e. Terakhir adalah menyimpan data ke format *csv*

2. *Preprocessing*

Preprocessing merupakan tahap yang dilakukan sebelum proses pengklasifikasian sentimen. Proses ini dilakukan untuk mengolah dataset agar siap digunakan untuk tahap selanjutnya.

Adapun urutan-urutan dalam melakukan *preprocessing* pada sebuah *tweet* :

- a. Menghilangkan URL.
- b. Menghilangkan *Emoticon*.
- c. Menghilangkan Simbol spesial dari *Twitter*. Simbol spesial yang dimaksud adalah seperti *hashtag*(#), *username*(@*username*) dan *retweet*(RT).

- d. Menghilangkan *ASCII number* dan *Punctuation*
- e. Tokenisasi, membagi seluruh kata-kata di dalam *tweet* menjadi sebuah token/term.
- f. *Case folding*, mengubah seluruh karakter di dalam *tweet* menjadi *lower case*.
- g. Mengubah kata tak baku menjadi baku
- h. Stemming, menghilangkan kata-kata yang berimbuhan. Hasil dari tahap ini adalah kata dasar.

Hasil dari *preprocessing* akan menghasilkan suatu *clean word* yaitu kata yang bebas dari *noisy text* yang tidak mempengaruhi dalam langkah-langkah pembangunan sistem selanjutnya.

3. Feature Extraction

Pada tahap *feature extraction* ini, dilakukan pembobotan kata pada data yang sudah dikumpulkan. Metode TF-IDF merupakan metode yang akan digunakan pada penelitian ini. Metode TF-IDF akan menentukan pembobotan kata berdasarkan frekuensi atau banyaknya kemunculan suatu kata dalam sebuah dokumen. Apabila presentase munculnya suatu kata semakin besar, maka nilainya akan semakin besar.

4. Klasifikasi Metode *K-Nearest Neighbor*

Metode ini dilakukan dengan memodifikasi nilai *k-values*. Langkah awal yang dilakukan adalah pemeringkatan hasil dari pembobotan,

kemudian dilakukan perhitungan nilai *k-values* baru atau biasa dan disebut dengan nilai *n*.

1.6.3 Metode Implementasi

Tahap implementasi merupakan penerapan dari metode analisis, seperti data collecting, *preprocessing*, pembobotan kata (*feature extraction*), dan klasifikasi metode *K-Nearest Neighbor*.

1.6.4 Metode Testing

Pada tahap ini, dilakukan pengujian *precision*, *recall*, *f-measure*, dan *accuracy* yang akan disajikan dalam *confusion matrix* untuk mengetahui rata-rata keberhasilan dari sistem.

1.7 Sistematika Penelitian

Sistematika penulisan pada penelitian ini adalah:

1. BAB I Pendahuluan

Bab ini peneliti menjelaskan latar belakang, rumusan masalah, batasan masalah, tujuan masalah, manfaat, metode, dan sistematika penelitian. sistematika penulisan.

2. BAB II Landasan Teori

Bab ini berisi gambaran mengenai analisis sentimen pada twitter dengan menggunakan metode *K-Nearest Neighbor*.

3. BAB III Metodologi Penelitian

Pada bab ini, peneliti akan memaparkan apa saja yang terdapat pada metodologi penelitian yaitu bahan penelitian, alat penelitian, metode penelitian, dan jadwal penelitian.

4. BAB IV Hasil dan Pembahasan

Bagian ini berisi analisis dari hasil pengolahan data dan pembahasan mengenai sentimen analisis yang dilakukan terhadap akun twitter milik provider internet Indihome.

5. BAB V Kesimpulan dan Saran

Kesimpulan merupakan uraian singkat hasil dari penelitian dan saran-saran yang diharapkan dapat bermanfaat bagi peneliti selanjutnya yang serupa.