

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada saat ini internet semakin berperan penting dalam kehidupan manusia (social,ekonomi,politik) Setiap tahunnya internet semakin dalam mempengaruhi kehidupan manusia. Indonesia juga memiliki kenaikan jumlah penggunaan internet pada tahun 2018 Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) memiliki data jumlah pengguna internet di Indonesia sebesar 171,1 juta naik sebesar 27,9 juta dari tahun 2017 yang berjumlah 143,2 juta pada tahun 2019-2020 pengguna internet menjadi 196,71 juta dari jumlah penduduk Indonesia 266,91 juta jadi 73,7% penduduk negara Indonesia menggunakan internet. Ada beragam alasan seseorang menggunakan internet diantaranya penggunaan sosial media menjadi sesuatu yang sering digunakan pada tahun 2019-2020 mencapai 51,5% dengan menggunakan (*Facebook, Twitter,Instagram*) dan penggunaan internet yang sering diakses selanjutnya adalah komunikasi lewat pesan mencapai 32,9% dan lain-lain [1].

Sosial media seperti *Twitter* adalah salah satu sosial media yang cukup populer di Indonesia. Pada tahun 2012 Indonesia sendiri menempati peringkat kelima dalam pengguna terbesar di *Twitter*[2]. Sosial media yang satu ini biasa digunakan untuk berbagi informasi antar pengguna atau pesan yang terdiri dari 140 karakter yang biasa disebut dengan kata *tweet* [3].

Twitter seringkali digunakan untuk mengungkapkan sesuatu emosi mengenai berbagai hal, baik itu dalam memuji maupun mencela. Emosi dalam diri manusia dapat dikategorikan menjadi lima emosi dasar seperti takut, marah, sedih, senang, dan cinta. Emosi cinta dan senang masuk ke dalam emosi positif sedangkan Emosi takut, marah, dan sedih masuk ke dalam emosi negatif [4].

Analisis mengenai pesan (*tweet*) dari pengguna *Twitter* disebut sebagai analisis opini dan sentiment (*opinion analysis and sentiment analysis*). Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan beropini pengguna terhadap sebuah objek atau masalah, apakah bersifat opini positif atau negatif [5]. Objek dalam penelitian disini adalah layanan jasa internet First Media.

First Media merupakan perusahaan layanan jasa *Internet Service Provider* (ISP) yang menyediakan layanan telekomunikasi dan multimedia yang mengusung konsep *TriplePlay*, yaitu layanan televisi berlangganan, layanan internet pita lebar dan komunikasi data. Melalui sosial media *Twitter* pengguna dapat memberikan opini seperti kritik serta saran untuk menilai kepuasan mereka kepada layanan internet dari First Media. Untuk dapat melakukan hal itu, dibutuhkannya sebuah sistem yang dapat menganalisis sentiment berdasarkan pesan (*tweet*) pengguna menggunakan *text mining* dengan klasifikasi sentimen.

Text mining adalah proses pengetahuan intensif dimana pengguna berinteraksi koleksi dokumen dari waktu ke waktu dengan menggunakan rangkaian alat analisis. *text mining* juga bisa diartikan dengan definisi menambang data berupa teks dimana sumber data yang didapatkan berasal dari dokumen. Klasifikasi sentimen pada penelitian ini menggunakan metode K-Nearest Neighbor (K-NN) dan *Naïve Bayes* untuk menilai suatu opini apakah termasuk dalam sentiment positif atau negatif, K-Nearest Neighbor sendiri adalah algoritma yang cukup efisien untuk klasifikasi teks, sifat dari K-Nearest Neighbor sendiri adalah *self-learning* dimana algoritma ini dapat mempelajari struktur data yang ada dan mengkategorikan dirinya sendiri [6]. Sehingga dengan melakukan penelitian ini diharapkan dapat mengetahui tingkat akurasi dan sentiment terhadap layanan internet First Media berdasarkan pesan (*tweet*) dari sosial media *Twitter*.

Penelitian mengenai analisis sentimen sudah pernah dilakukan pada layanan *E-commerce* dengan menggunakan metode K-Nearest Neighbor. Pada penelitian ini data yang

digunakan didapat dengan *crawling* pada *Twitter*. Setelah itu data diproses pada tahap *Text preprocessing* bagian dimana mengurangi atribut yang kurang berpengaruh terhadap proses klasifikasi dengan beberapa tahapan yaitu *case floding*, *cleaning*, *convert emoticon*, *tokenizing*, *stopword removal*, dan *stemming*. Akurasi yang didapat dari penelitian ini sebesar 88.11% dengan nilai $k = 1$ [7]. Dalam penelitian [8] menggunakan *Naïve Bayes* analisis sentiment mengenai opini public terhadap COVID-19, menghasilkan tingkat akurasi sebesar 79,20% dikarenakan penelitian sebelumnya tingkat akurasi yang cukup tinggi maka untuk penelitian tugas akhir ini akan menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes*.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah adalah sebagai berikut :

1. Apa pendapat pelanggan First Media pada *Twitter* terhadap layanan internet First Media ?
2. Berapa tingkat akurasi metode *K-Nearest Neighbor* dan *Naïve Bayes* dalam melakukan klasifikasi analisis sentimen pada media sosial *Twitter* khususnya sentiment analisis layanan internet First Media ?

1.3 Batasan Masalah

Batasan-batasan masalah pada penelitian ini antara lain :

1. Klasifikasi sentimn dibagi menjadi tiga, yaitu positif ,netral dan negatif.
2. Menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes*.
3. Untuk mengambil data pada *Twitter* menggunakan *Twitter API*.
4. Menggunakan fitur pembobotan kata TF-IDF pada algoritma K-NN.
5. Pembuatan system menggunakan bahasa pemrograman *Python*.
6. *Tweet* yang diambil untuk dianalisis hanya *tweet* berbahasa Indonesia.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini yaitu untuk :

1. Mengetahui pendapat pelanggan First Media mengenai layanan internet First Media dari media sosial *Twitter*.
2. Mengetahui tingkat akurasi dari metode *K-Nearest Neighbor* dan *Naïve Bayes* dalam pengklasifikasian pendapat pelanggan First Media pada *Twitter* terhadap layanan internet First Media.

1.5 Manfaat Penelitian

Manfaat yang diberikan dari penelitian ini sebagai berikut :

1. Membantu memberikan *feedback* mengenai layanan internet First Media.
2. Membantu menganalisis sentimen pada *Twitter* untuk layanan internet First Media dengan metode *K-Nearest Neighbor* dan *Naïve Bayes*.
3. Menambah referensi bagi penelitian berikutnya yang relevan untuk kasus analisis sentimen *Twitter*.

1.6 Metode Penelitian

Metode yang diharapkan dari penelitian ini berupa suatu sistem yang mampu mengelompokkan *tweet* berisikan opini positif, netral dan negatif pada *Twitter*.

1. Metode Pengumpulan Data

Peneliti memperoleh data dilakukan dengan studi kepustakaan dan studi literatur, yaitu dengan mencari buku-buku, jurnal, skripsi, tutorial dan berbagai informasi yang berkaitan sebagai referensi dan melakukan observasi dengan cara *crawling* data dari *Twitter*.

2. Metode Analisis

Pada tahap ini dilakukan analisis terhadap studi literatur untuk mendapatkan pemahaman mengenai metode yang akan digunakan, yaitu pada metode *K-*

Nearest Neighbor (K-NN) dan *Naïve Bayes Classifier* (NBC) untuk menyelesaikan masalah dalam analisis sentimen.

3. Metode Perancangan Desain Sistem

Ditahap ini dilakukan perancangan berdasarkan spesifikasi yang sudah ditentukan sebelumnya. Perancangan terfokus pada pengumpulan data, algoritma yang digunakan dan *interface* atau antarmuka untuk pengujian. Proses perancangan dilakukan berdasarkan hasil yang didapat dari analisis studi literatur yang diperoleh.

1.7 Sistematika Penulisan

Pada bagian ini terdiri dari pokok-pokok pembahasan dan permasalahan yang dibagi menjadi beberapa bagian, yaitu :

BAB I PENDAHULUAN

Bab ini membahas tentang gambaran umum penelitian dari latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, luaran, Batasan masalah, metode penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menyajikan tentang penjelasan teori-teori yang bersumber dari landasan teori yang berkaitan dengan sentimen, *text-processing* dalam pengambilan data *Twitter* dan algoritma *K-Nearest Neighbor* dan *Naïve Bayes*.

BAB III METODOLOGI PENELITIAN

Bab ini membahas tentang Langkah-langkah yang dilaksanakan dalam proses penelitian berkaitan dengan pengumpulan data dan metode yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisikan tentang penjabaran hasil penelitian, nilai akurasi yang didapat dari algoritma yang digunakan dan berbagai macam analisis hasil dari simulasi sistem yang telah dibuat.

BAB V KESIMPULAN

Bab ini berisikan tentang uraian kesimpulan dan saran berdasarkan pembahasan pada bab-bab sebelumnya

