

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring dengan kemajuan zaman, media massa yang digunakan untuk mencari informasi oleh masyarakat juga mengalami kemajuan yang semakin pesat, terutama teknologi internet. Salah satu bagian dari teknologi internet yang banyak dimanfaatkan oleh masyarakat yaitu penggunaan media sosial. Berbagai macam jenis media sosial contohnya Instagram, Facebook, Twitter, WhatsApp, dan lain-lain yang dimanfaatkan masyarakat untuk mengutarakan apa yang mereka rasakan, pendapat, pengetahuan, bahkan hal lain yang ingin mereka utarakan (Nugroho, 2018). Contoh jenis media sosial yang populer dan ramai digunakan yaitu Twitter.

Menurut Penelitian PeerReach bahwa Indonesia ialah negara pengguna twitter teraktif yang menempati posisi ke 3 di dunia. Hal ini membuktikan bahwa masyarakat di negara Indonesia ini merupakan pengguna twitter yang masuk ke dalam kategori pengguna yang aktif di dunia (Tarigan, 2013). Twitter merupakan suatu jenis sosial media yang menyediakan berbagai macam fitur. Salah satunya yaitu fitur “*tweet*”, fitur tersebut digunakan oleh pengguna untuk membuat status di twitter. *Tweet* tersebut dapat dibuat dengan batas maksimal 140 karakter. Dengan adanya fitur “*tweet*”, pengguna dapat mengungkapkan suatu pesan singkat yang mereka rasakan dan ingin membagikan serta dibaca oleh publik. (Kotwal *et al.*, 2016). Tidak jarang masyarakat mengakses berita-berita terkini yang ramai dibahas di twitter atau yang biasa disebut dengan istilah “*trending topic*”. Tentunya untuk mengakses twitter ini membutuhkan akses internet.

Akses internet yang digunakan dapat diperoleh dari penyedia jasa untuk mengakses layanan internet atau *Internet Service Provider (ISP)*. *Internet Service Provider* ialah suatu perusahaan yang mana memiliki jaringan cukup luas yang akan memberikan sebuah penawaran kepada *customer* seperti akses internet dengan kecepatan yang tidak diragukan alias tinggi, internet transit, dan

sebagainya. Salah satu perusahaan yang menyediakan layanan internet atau *Internet Service Provider* yang populer di kalangan masyarakat yaitu indihome.

Indihome ialah sebuah produk layanan dari PT Telekomunikasi Indonesia yang menyediakan paket jasa layanan internet dengan biaya atau tarif yang sudah ditentukan. Indihome ini merupakan penyedia *Internet Service Provider* yang sering digunakan oleh masyarakat daripada penyedia jasa layanan internet lainnya. Setiap produk yang ditawarkan di masyarakat tentunya memiliki kekurangan dan kelebihan masing-masing, begitupun juga indihome. Banyak masyarakat yang menyampaikan pendapat mereka tentang kekurangan dan juga kelebihan selama menggunakan indihome baik secara langsung maupun di sosial media.

Tanggapan atau umpan balik yang diberikan oleh masyarakat tentang pelayanan indihome ini merupakan aset penting bagi perusahaan untuk menentukan kualitas produk yang mereka tawarkan (Ahmad, 2017). Dengan demikian, untuk mengetahui bagaimana sentimen yang diberikan oleh masyarakat dapat dilakukan dengan menerapkan teknik *Sentiment Analysis*.

Sentimen Analysis merupakan proses komputasi untuk mengesktrak, memahami, dan juga mengolah suatu data yang tidak terstruktur menjadi data terstruktur. Dengan tujuan untuk menghasilkan suatu informasi sentimen yang terdapat pada pendapat, emosi, atau komentar seseorang dari dataset yang tidak terstruktur. *Sentiment analysis* memberikan dampak dan manfaat yang sangat besar dalam berbagai bidang, sehingga menyebabkan *sentiment analysis* ini berkembang dengan pesat dan banyak digunakan (Buntoro, 2017). Dalam melakukan analisis sentiment terdapat tahap ekstraksi fitur (*fitur extraction*).

Ekstraksi fitur merupakan suatu tahap untuk memproses sebuah kata yang menjelaskan suatu sentimen yang terdapat pada *dataset* untuk diekstraksi menjadi sebuah fitur atau aspek (Asghar *et al.*, 2014). Salah satu fitur yang paling sering digunakan ialah TF-IDF (*Term Frequency-Inverse Document Frequency*). Fitur TF-IDF mempunyai keunggulan dibandingkan dengan fitur lainnya yaitu pengimplementasinya yang mudah. Akan tetapi, kekurangan dari fitur TF-IDF ini adalah tidak dapat memproses relasi semantik antar kata sehingga fitur ini

menganggap setiap kata memiliki konteks yang berbeda (Paltoglou *et al.*, 2010). Fitur yang menggunakan relasi semantik dapat dilakukan dengan metode *Word2Vec*.

Word2Vec merupakan suatu metode yang merepresentasikan suatu kata pada ruang vektor dengan dimensi yang tinggi. Selain itu, metode *word2vec* juga dapat mengetahui hubungan semantik antar kata dengan cara menghitung *cosine similarity* antar kata melalui nilai vektor (Kurniawan, 2020). Setelah dilakukan ekstraksi fitur, maka akan memasuki proses klasifikasi sentimen.

Penelitian ini, penulis menerapkan metode klasifikasi berupa *Naïve Bayes Classifier*. Metode *Naïve Bayes Classifier* ialah jenis metode klasifikasi dengan menggunakan algoritma sederhana akan tetapi mempunyai kecepatan serta nilai akurasi tinggi (Rish, 2001). Metode *Naïve Bayes Classifier* adalah suatu metode yang berdasar pada teori *Bayes* dengan asumsi yang kuat, dengan efek nilai atribut tidak bergantung pada kelas dan juga atribut lain. Teori *Bayes* merupakan suatu teori yang membahas klasifikasi untuk meramalkan atribut kelas suatu anggota probabilitas menurut data yang disediakan (Fauzi *et al.*, 2018).

Penelitian sebelumnya yang dilakukan oleh Beakcheol Jang (2019), telah dilakukan penelitian yang berjudul *Word2Vec Convolutional Neural Networks For Classification Of News Articles And Tweets*. Pada penelitian tersebut menjelaskan bahwa penerapan *word2vec* yang membahas tentang hubungan semantik antar kata secara signifikan mampu meningkatkan kinerja model klasifikasi. Selain itu, penelitian sebelumnya yang dilakukan oleh Susanti Fransiska (2020) dengan judul penelitian *Sentiment Analysis of Government Policy on Corona Case Using Naïve Bayes Algorithm*. Pada penelitian tersebut menggunakan seleksi fitur TF-IDF yang menghasilkan nilai akurasi sebesar 81% dengan nilai presisi sebesar 78%, *Recall* sebesar 91%, dan juga *f1-Score* sebesar 84%.

Mengacu pada latar belakang yang telah dipaparkan, penulis akan meneliti tentang review masyarakat terhadap pelayanan indihome berdasarkan data twitter dengan mengimplementasikan metode klasifikasi *Naïve Bayes Classifier* dan melakukan perbandingan ekstraksi fitur antara TF-IDF dan *Word2Vec*. Maka

judul penelitian ini ialah “Pengaruh Ekstraksi Fitur terhadap Analisis Sentiment pada Data Review Pelayanan Indihome Berbasis Naïve Bayes”.

1.2 Rumusan Masalah

Rumusan masalah berdasarkan uraian latar belakang diatas ialah:

1. Berapa jumlah sentiment yang bernilai positif dan negatif pada klasifikasi analisis sentimen layanan indihome di media sosial twitter?
2. Berapa perbandingan tingkat akurasi yang diperoleh antara menggunakan ekstraksi fitur TF-IDF dan *Word2Vec* pada analisis sentimen layanan indihome di media sosial twitter?

1.3 Tujuan

Tujuan diadakannya penelitian ini yaitu:

1. Mengetahui jumlah sentiment yang bernilai positif dan negatif pada analisis sentimen layanan indihome di twitter.
2. Mengetahui perbandingan tingkat akurasi mana yang lebih besar antara ekstraksi fitur TF-IDF dan *Word2Vec* dalam melakukan analisis sentimen layanan indihome di twitter.

1.4 Manfaat

Hasil penelitian ini diharapkan bermanfaat bagi peneliti lain dalam bidang analisis sentiment secara umum dengan menggunakan metode *naïve bayes* serta ekstraksi fitur TF-IDF dan *Word2Vec*.

1.5 Batasan Masalah

Penelitian ini memiliki beberapa batasan masalah sebagai berikut:

1. Data yang didapatkan sebanyak 655 data pada bulan Juli 2021.
2. Proses pencarian data dengan menggunakan kata kunci “indihome”.

3. Penelitian ini memerlukan API (*Application Programming Interface*) Twitter penulis yang digunakan untuk proses crawling, yang didapatkan dari Twitter Development.
4. Hasil klasifikasi akhir berupa 2 kelas yaitu sentiment positif dan juga negatif.
5. Menggunakan *K-fold cross validation* dengan $K=2, 3, 5, 7,$ dan 10 .
6. Metode yang diimplementasikan pada penelitian ini yakni metode *Naïve Bayes Classifier* serta menerapkan fitur *Word2Vec* serta TF-IDF untuk pembobotan kata.
7. Menggunakan bahasa pemrograman *python* dan *software* yang digunakan berupa *Jupyter Notebook*.

