

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dengan semakin berkembang pesatnya teknologi informasi di Indonesia. Tidak heran jika pengguna media sosial di Indonesia semakin meningkat, meningkatnya pengguna media sosial di Indonesia karena semakin lengkapnya fasilitas yang disediakan oleh penyedia jasa layanan komunikasi sehingga hampir seluruh golongan masyarakat di Indonesia dapat menikmati media sosial. Sesuai dengan riset yang dilakukan oleh APJII (Asosiasi Penyedia Jasa Internet Indonesia) di tahun 2018, pengguna internet di Indonesia mencapai 64,8% (171,17 juta jiwa). Dari jumlah pengguna internet tersebut, tercatat 90,7% adalah penggunaan internet untuk mengakses sosial media. Lima media sosial dengan jumlah akses yang terbanyak adalah Facebook, Instagram, YouTube, Twitter, dan LinkedIn. Sosial media YouTube tercatat memiliki 15,1% pengguna di Indonesia menurut data statistik APJII (Asosiasi Penyedia Jasa Internet Indonesia) di tahun 2018 [1]. YouTube memberikan fasilitas bagi pengguna YouTube dapat meng-upload video agar dapat diakses oleh pengguna yang lain.

YouTube juga memberikan fasilitas untuk pengguna berupa Kategori seperti Berita, Game, Musik dan Trending. Salah satu fasilitas yang memudahkan pengguna dalam mencari konten video yang sedang ingin ditonton oleh kebanyakan pengguna YouTube adalah Kategori Trending, video yang masuk dalam Kategori Trending terdapat jumlah *viewers* dan komentar yang tidak sedikit. Dan tidak

sedikit juga konten video yang mengisi Kategori Trending di YouTube dari Channel YouTuber Indonesia. Tidak dapat dipungkiri bahwa video yang masuk dalam Kategori Trending dengan jumlah *viewers* dan komentar yang tidak sedikit, terdapat komentar seperti komentar positif, netral, dan negatif.

Pada penelitian ini, akan melakukan analisis sentimen pada kolom komentar video yang sedang Trending di YouTube. Dengan menganalisis komentar pada video yang sedang Trending, dapat diketahui kategori komentar positif atau negatif yang sering dilontarkan oleh *viewers* pada kolom komentar video yang sedang Trending di YouTube. Melihat apa yang sudah diuraikan diatas, atas dasar tersebut penelitian ini mengambil judul **“Analisis Sentimen Terhadap Komentar Video Trending YouTube Menggunakan Metode Naive Bayes Classifier”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan seperti diatas, maka permasalahan yang dapat dirumuskan yaitu: Bagaimana hasil klasifikasi sentimen positif dan negatif menggunakan kernel *multinomial* pada metode Naive Bayes dan bagaimana akurasi yang akan diperoleh dari kernel *multinomial* pada metode Naive Bayes dengan menerapkan kombinasi pembobotan kata untuk melakukan klasifikasi sentimen komentar pada video Trending YouTube?

1.3 Batasan Masalah

Berdasarkan rumusan masalah, beberapa batasan masalah dalam analisis ini adalah sebagai berikut:

1. Video yang dianalisis berasal dari YouTuber Indonesia dan menggunakan bahasa Indonesia.
2. Komentar dikumpulkan dari video yang masuk Trending YouTube.
3. Komentar yang dianalisis maksimal 2000 komentar.
4. Tidak ada pengecualian pada genre video yang akan di analisis.
5. Metode Naïve Bayes digunakan untuk memprediksikan sentiment positif dan negatif.

1.4 Manfaat dan Tujuan Penelitian

1.4.1 Tujuan Penelitian

Penelitian ini bertujuan untuk mengetahui hasil klasifikasi komentar positif dan negatif menggunakan kernel *multinomial* pada Naïve Bayes dan mengetahui berapa akurasi yang diperoleh untuk menganalisis sentimen positif dan negatif pada video Trending YouTube jika menggunakan kernel *multinomial* pada metode Naive Bayes.

1.4.2 Manfaat Penelitian

Beberapa manfaat dari analisis ini yaitu:

1. Analisis sentimen yang dilakukan pada penelitian ini dapat dijadikan referensi, serta dapat dijadikan sebagai acuan atau pengembangan bagi para peneliti yang lainnya dalam studi kasus penelitian sejenis.

2. Dengan dilakukannya analisis sentimen pada penelitian ini, dapat diketahui komentar positif atau negatif yang lebih mendominasi pada video yang masuk ke dalam Trending YouTube.
3. Dapat diketahui nilai akurasi kernel *multinomial* pada algoritma Naïve Bayes dalam memprediksi komentar positif dan negatif pada video yang masuk ke dalam Trending YouTube.

1.5 Metode Penelitian

Metode penelitian yang dilakukan adalah sebagai berikut:

1.5.1 Metode Pengumpulan Data

Metode pengumpulan data untuk memenuhi kebutuhan analisis pada penelitian ini, yaitu:

1.5.1.1 Studi Pustaka

Peneliti mengumpulkan jurnal atau literatur yang berkaitan dengan *text mining*, metode *term weighting* seperti TF-IDF, kernel pada algoritma *Naïve Bayes Classifier*.

1.5.1.2 Text Mining

Dataset komentar diambil menggunakan YouTube Data API V3 pada video Trending YouTube.

1.5.2 Metode Analisis

1.5.2.1 Text Preprocessing

Pada tahap ini berfungsi untuk membersihkan data supaya lebih terstruktur dan format data sesuai dengan tugas data mining, pada *text preprocessing* terdapat tahapan proses seperti *case folding*, *tokenizing*, *filtering* dan *stemming*. Dengan dilakukannya preprocessing pada data dapat memudahkan dalam pembobotan kata dan menghasilkan akurasi dan hasil yang berkualitas dari kinerja data mining.

1.5.2.2 Pembobotan Kata

Pembobotan kata berfungsi untuk memberikan bobot/nilai pada term yang ada pada data, pembobotan yang diberikan pada term data tergantung pada metode yang akan digunakan. Terdapat beberapa macam algoritma yang ada pada pembobotan kata, pada penelitian ini akan menggunakan salah satu algoritma pembobotan kata yaitu TF-IDF.

1.5.2.3 Klasifikasi Naïve Bayes

Pada analisis sentimen, klasifikasi merupakan tahap yang sangat penting. Dimana pada tahap ini data yang sudah melalui proses text preprocessing dan pembobotan kata menggunakan TF-IDF akan di klasifikasi menggunakan metode *Naive Bayes*, yang menghasilkan informasi data termasuk dalam komentar positif atau negatif.

1.5.3 Metode Pengujian

Untuk menguji kinerja dari model klasifikasi, metode yang digunakan adalah *confusion matrix*.

1.5.3.1 Akurasi

Untuk mengetahui seberapa akurat model dalam mengklasifikasi dengan benar. Untuk memastikan akurasi model, maka peneliti menggunakan *K-Fold Cross Validation*.

1.5.3.2 Presisi

Presisi berfungsi untuk mengukur tingkat keakuratan antara data dengan hasil prediksi yang diberikan oleh model.

1.5.3.3 Recall

Pengujian ini dapat menggambarkan seberapa besar, keberhasilan model dalam menemukan kembali sebuah informasi.

1.5.3.4 F1 Skor

Sering terjadi komplikasi antara presisi dan *recall* yang disebabkan adanya *trade off* di antara keduanya. F1 skor merupakan *harmonic mean* dari presisi dan *recall*. Untuk menghindari dilema tersebut, maka skor yang digunakan adalah f1 skor.

1.6 Sistematika Penulisan

Peneliti membagi beberapa bab untuk mempermudah penulis dalam menyusun dan mempermudah bagi pembaca untuk memahami, yang mana tiap-tiap bab terdiri dari beberapa sub bab yang merupakan penjelasan dari bab sebelumnya. Berikut uraian singkat dari setiap bab:

BAB I PENDAHULUAN

Pada bab ini berisi uraian tentang skripsi yang akan dibuat yang berisi Latar Belakang Masalah, Rumusan Masalah dan Batasan Masalah, Maksud dan Tujuan, Manfaat Penelitian dan Sistematika Penulisan.

BAB II LANDASAN TEORI

Pada bab ini menjelaskan tentang landasan teori dan kajian pustaka dari penelitian yang memiliki keterkaitan dengan penelitian ini. Kajian pustaka dapat memperkuat dasar dan alasan dilakukannya penelitian ini. Selain kajian pustaka, pada bab ini juga menjelaskan tentang teori - teori yang terkait dengan Term Frequency Inverse Document Frequency (TF-IDF), teori - teori yang berkaitan dengan klasifikasi kata sentimen dan metode Naive Bayes.

BAB III METODE PENELITIAN

Pada bab ini akan diuraikan mengenai tahapan dalam pengumpulan data, analisis data, pembobotan kata, sampai tahap klasifikasi.

BAB IV IMPLEMENT DAN PEMBAHASAN

Pada bab keempat ini menjelaskan tentang implementasi analisis metode *Naive Bayes* untuk klasifikasi komentar positif dan negatif menggunakan metode Naive Bayes serta *Confusion Matrix* dan *K-Fold Validation* dalam menghitung nilai akurasinya.

BAB V PENUTUP

Pada bab terakhir ini berupa kesimpulan dari penelitian dan juga tentang segala kekurangan dan kelebihan dalam menerapkan metode Naïve Bayes pada penelitian ini.

