

BAB III

LANDASAN TEORI

Pada bab ini akan membahas mengenai bagian-bagian yang menjadi landasan dalam melakukan analisis sentimen terhadap opini publik mengenai tayangan media televisi.

3.1 Analisis Sentimen

Analisis sentimen merupakan sebuah ilmu yang menganalisis pendapat, sifat, dan emosi seseorang terhadap suatu produk tertentu. Selain terhadap suatu produk juga dapat terhadap pelayanan, sebuah organisasi, individu, permasalahan tertentu, sebuah even tertentu dan informasi yang menyebar secara luas dengan cepat yang sedang dibicarakan oleh banyak masyarakat [8]. Analisis sentimen juga diartikan sebagai proses mengklasifikasikan dokumen menjadi dua bagian yaitu sentimen yang positif dan sentimen yang negatif. Tujuan dilakukan proses analisis sentimen ini adalah untuk mengetahui sentiment publik terhadap suatu permasalahan dengan menggunakan pendekatan dalam machine learning (Putranti,2014). Analisis sentimen merupakan tipologi area dari Ilmu Natural Language Processing (NLP) dengan penelitian mulai dari klasifikasi tingkat dokumen hingga mempelajari polatiras kata dan frasa [8].

3.2 Data Mining

Data *mining* adalah ilmu yang bergerak pada bidang komputer, banyak hal yang diterapkan seperti bidang ilmu (*artificial intelligence*, database, statistik, pemodelan matematika dan pengolahan citra). Dengan menerapkan hal ini membuat ilmu penerapan ini menjadi makin luas. Data mining merupakan suatu kegiatan yang mengekstraksi data-data yang berukuran atau berjumlah yang cukup besar, selanjutnya akan

menghasilkan informasi yang nantinya akan berguna untuk pengembangan selanjutnya. Nama lain dari data mining juga adalah Knowledge Discovery in Databases (KDD) [9].

3.3 *Twitter*

Twitter merupakan sebuah aplikasi pengirim pesan dimana aplikasi ini memungkinkan *User* dapat mengirimkan informasi dengan cepat berupa pesan antara satu *User* dan *User* lainnya yang saling berhubungan dalam arti memiliki relasi pertemanan. Namun seiring berjalannya waktu, saat ini *twitter* telah berkembang dan menambah banyak fitur yang menjadikan *twitter* ini menjadi jejaring sosial media yang dapat melakukan banyak hal seperti dapat membagikan informasi status diakses menggunakan Internet.

Twitter didirikan oleh Jack Dorsey, Evan Williams dan Biz Stone. Pada saat tanggal 15 Juli 2006 pembuatan proyek *twitter* ini dilakukan. Kemudian saat April 2007 *twitter* secara resmi menjadi sebuah perusahaan, dengan menggunakan nama *twitter*.

Ada beberapa fitur yang dapat diakses pada aplikasi *twitter*, yaitu:

- a. Timeline: merupakan fitur yang menampilkan tweet dari *User twitter* yang di ikuti oleh pengguna akun, termasuk menampilkan tweet yang dibuat pemilik akun, dimana pada Timeline ini menampilkan informasi tweet yang terbaru.
- b. Direct Message (DM) yaitu fitur yang digunakan untuk saling mengirim informasi berupa pesan antar pengguna yang keamanan privasinya lebih terjaga. DM hanya bisa dilakukan oleh pihak yang diikuti.
- c. Trending topics merupakan daftar topik-topik yang sedang hangat diperbincangkan pada kalangan pengguna *twitter*.
- d. Tweet adalah fitur yang memberikan informasi yang terdiri dari 140 karakter. Tweet berisi berita terbaru ("apa yang sedang terjadi") yang berkaitan dengan hal - hal yang pemilik akun sukai.

- e. Reply tweet atau response tweet (RT) adalah balasan atas tweet.
- f. Retweet adalah menyalin seluruh isi tweet dari akun lain.
- g. Follow adalah sebuah fitur untuk mengikuti akun seorang pengguna.
- h. Follower adalah fitur yang memberikan informasi yang pengguna akun mengikuti akun seseorang.
- i. Mention (@) fitur ini digunakan untuk menyebut username akun pengguna yang akan diajak untuk saling berkomunikasi.
- j. Hastags adalah fitur yang digunakan untuk menandai kata kunci untuk topik diskusi atau informasi yang dibagikan agar mudah dicari[10].

3.4 Opini Publik

Arti dari opini yaitu tanggapan atau pendapat yang merupakan suatu jawaban yang terbuka terhadap suatu persoalan atau isu yang telah beredar. Menurut Cultip dan Center opini merupakan suatu ekspresi terhadap sikap mengenai suatu hal yang bersifat kontroversial. Opini yang muncul sebagai hasil pembicaraan terhadap suatu permasalahan yang kontroversial. Hal ini menimbulkan pendapat yang berbeda-beda.

Opini merupakan tanggapan atau pendapat seseorang mengenai suatu hal yang diturunkan dan akan menimbulkan perasaan, pikiran dan kesediaannya terhadap sesuatu yang telah terjadi. Unsur-unsur yang merupakan molekul dari sebuah opini, yaitu *belief* (kepercayaan tentang sesuatu), *attitude*, dan persepsi.

Bernard Berelson dalam tulisannya yang berjudul “*Communication and Public Opinion*” mengatakan bahwa pendapat diartikan sebagai “*people response* atau jawaban masyarakat terhadap isu atau hal-hal yang bersifat politis dan sosial dan memerlukan perhatian umum. [19]

3.5 Metode Naïve Bayes Classifier

Klasifikasi naïve bayes adalah metode klasifikasi yang mencari nilai probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi independen yang tinggi. Keuntungan penggunaan klasifikasi *Naive Bayes* adalah metode ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Metode ini meliputi dua tahap dalam proses klasifikasi teks, yaitu pada tahap pertama, tahap pelatihan akan dilakukan proses terhadap sampel data yang nantinya data tersebut akan menjadi representasi data. Kemudian menentukan nilai probabilitas *prior* dari setiap kategori berdasarkan sampel data yang ada. Lalu pada tahap selanjutnya yaitu tahap klasifikasi. Pada tahap ini akan menentukan nilai kategori dari sampel data yang ada berdasarkan *term* yang muncul pada data yang telah diklasifikasikan [11].

Metode ini melihat setiap tulisan atau teks dengan sifat positif, negatif, dan netral. Pada metode ini memiliki tingkat akurasi kecepatan yang tinggi dan dapat dimasukkan ke dalam database. Berikut rumus dasar yang dipakai dalam metode *naïve bayes classifier* :

$$P(C_i|X) = \frac{P(X|C_i) * P(C_i)}{P(X)}$$

Keterangan :

$P(X|C_i)$ = Nilai Probabilitas X terjadi jika C_i sudah terjadi berdasarkan data training.

$P(C_i)$ = Nilai probabilitas C_i dalam data dengan sifat independen terhadap nilai X.

X = X merupakan kumpulan atribut.

$P(C_i|X)$ = Merupakan probabilitas C_i terjadi jika X sudah terjadi.

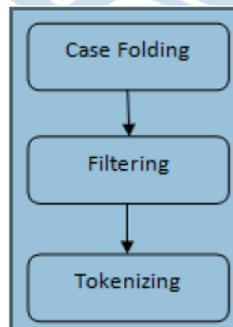
$P(X)$ = Probalitas dari X.

3.6 Pengumpulan Data

Tahap ini adalah tahap pengumpulan data dimana peneliti melakukan pencarian dan pengumpulan data yang berkaitan dengan tes CPNS (Calon Pegawai Negeri Sipil). Pengumpulan data dengan menggunakan Salah satu media sosial yaitu *Twitter* diimplementasikan ke dalam bahasa pemrograman python aplikasi python 3.8. Data ini berdasarkan hasil tweet dengan metode crawling terkait dengan tes calon pegawai negeri sipil. Untuk memperbanyak data dalam penelitian ini juga menggunakan hastag untuk mendukung dalam proses pencarian data dan dilanjutkan pada tahap preprocessing atau cleaning.

3.7 Preprocessing atau cleaning

Pada tahap ini merupakan proses pembersihan data dimana dataset yang didapat sebelum digunakan akan dilakukan *preprocessing data* sehingga dataset telah bersih akan siap digunakan pada proses selanjutnya. Berikut merupakan alur *preprocessing data*.



Tahap yang pertama adalah *preprocessing data* atau *case folding*, yaitu pada proses ini akan mengubah semua huruf kapital menjadi huruf kecil. Kemudian pada Tahap kedua *filtering* berupa tahap yang membuang semua karakter selain karakter string serta menghapus beberapa karakteristik dari data yang didapat dari *twitter*, seperti *@username*, *#hashtag*, atau kata yang bersifat perulangan. Tahap filtering ini juga menghapus *stopword*. Hal ini berfungsi untuk mengurangi load saat melakukan testing dataset.

Kemudian dilanjutkan dengan tahap tokenizing, yaitu melakukan proses pemisahan berdasarkan perkata. Langkah ini dapat dilakukan dengan memberi tanda pada karakter spasi sebagai pembatas[11].

3.8 Python 3.8

Python adalah bahasa pemrograman yang diinterpretasikan, sehingga dengan penggunaan interpreter baris perintah perancangan yang memiliki tingkat keterbacaan kode. Pengguna dapat dengan mudah memeriksa bagaimana fungsi bekerja. Penggunaan Python cukup mudah, bahwa seseorang yang lulus kursus matematika dasar akan menemukan alat seperti 'variabel' dan 'Fungsi' dilengkapi sintaks kode yang begitu jelas dikarenakan kode-kode yang nantinya akan di compile menjadi *byte code* lalu di eksekusi sehingga menjadikan bahasa pemrograman[12].

3.9 WEKA

Weka adalah sebuah aplikasi yang pertama kali dikembangkan oleh Universitas Waikato di Selandia Baru, yang kemudian menjadi bagian dari Pentaho. Weka salah satu aplikasi untuk *data mining* yang berbasis *Java*. Weka sendiri terdiri dari kumpulan algoritma *machine learning*, yang digunakan untuk melakukan generalisasi, formulasi dari kumpulan *data sampling*. Weka mempunyai beragam *tools* untuk keperluan pengolahan data, yaitu *pre-processing*, *classification*, *regression*, *clustering*, *association rules*, dan *visualization*.

3.10 Confusion Matrix

Confusion matrix merupakan metode yang digunakan untuk mengukur atau melakukan perhitungan akurasi pada konsep *data mining*. *Confusion matrix* terdiri dari baris data uji yang diprediksi benar dan tidak benar suatu data oleh model klasifikasi [21]. Berikut gambar dibawah ini sebuah *confusion matrix* untuk pengklasifikasikan ke dalam dua kelas.

	Predicted Positives	Predicted Negatives
Actual Positives instances	Number of True Positives instances (TP)	Number of False Negatives instances (FN)
Actual Negatives instances	Number of False Positives instances (FP)	Number of True Negatives instances (TN)

Gambar 3.1 *confusion matrix*

Berdasarkan gambar *confusion matrix* diatas :

True positives (TP) merupakan jumlah data positif yang diklasifikasikan sebagai nilai positif.

False positives (FP) merupakan jumlah data negatif yang diklasifikasikan sebagai nilai positif.

False negatives (FN) merupakan jumlah data positif yang diklasifikasikan sebagai nilai positif.

True negatives (TN) merupakan jumlah data positif yang diklasifikasikan sebagai nilai negatif.

Nilai yang dihasilkan melalui metode *confusion matrix* adalah berupa evaluasi sebagai berikut :

a. Akurasi

Merupakan presentase jumlah data yang dilakukan pada klasifikasi atau prediksi secara benar oleh algoritma

$$\text{Rumus : } Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

b. Presisi

Nilai dari ketepatan dari metode yang digunakan dalam klasifikasi. Nilai tersebut menunjukkan banyaknya data yang dapat terklasifikasi di kelas yang benar dalam beberapa pengujian.

$$\text{Rumus : } Presisi = TP / (TP + FP)$$

c. Recall

Nilai yang dapat mengukur hasil berapa presentase data yang terklasifikasikan dengan benar, adapun rumus yang digunakan untuk menghitung nilai recall adalah :

$$\text{Recall} = TP / (TP + FN)$$

3.11 *N-gram*

N-gram merupakan sekumpulan kata yang diambil dari sebuah paragraf dan kalimat. *N-gram* digunakan untuk mengambil kata perkata dari sebuah kontinuitas dibaca dari teks sumber hingga akhir dari dokumen. *N-gram* dibedakan berdasarkan jumlah potongan karakter sebesar n . Untuk membantu dalam mengambil kata perkata yang berupa karakter huruf tersebut, maka dilakukan padding dengan blank diawal dan diakhir suatu kata.

