

BAB 3

LANDASAN TEORI

3.1. Data Mining

Data mining adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Secara teknis, *data mining* disebut sebagai proses untuk menemukan pola dari ratusan atau ribuan *field* dari sebuah relasional *database* yang besar (Ginanjar & Lubis, 2012).

Istilah *data mining* dan *knowledge discovery in databases* (KDD) seringkali digunakan untuk menjelaskan proses penggalian informasi yang tersembunyi dalam suatu *database* yang besar, kedua istilah tersebut memiliki konsep berkaitan satu sama lain yaitu salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Ada pun proses-prosesnya diuraikan sebagai berikut (R & H., 2008) :

- a. *Data Selection* : Seleksi data dilakukan pada sekumpulan data operasional. Seleksi data perlu dilakukan sebelum tahap penggalian informasi dimulai. Data hasil seleksi tersebut akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, dan terpisah dari data operasional.

- b. *Pre-processing / Cleaning* : Proses *cleaning* atau pembersihan pada data menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak.
- c. *Transformation* : Proses transformasi dilakukan pada data yang telah dipilih melalui proses *coding*, sehingga data tersebut nantinya sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif yaitu sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.
- d. *Data mining* : Proses mencari pola atau informasi yang menarik didalam data yang sudah diseleksi atau dipilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* bermacam-macam. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
- e. *Interpretation/ Evaluation* : Tahap ini mencakup pemeriksaan atau perbandingan apakah pola atau informasi yang ditemukan bertolak belakang dengan fakta atau hipotesa yang ada sebelumnya. Pola informasi yang dihasilkan dari proses *data*

mining sebelumnya, perlu ditampilkan dalam bentuk yang mudah agar dimengerti oleh pihak yang berkepentingan.

Adapun proses yang diperlukan dalam tahap *preprocessing* yaitu *Stemming*. *Stemming* dilakukan untuk mengenali setiap kata yang dimasukkan, proses pemetaan dan penguraian berbagai bentuk dari suatu kata menjadi bentuk kata dasarnya atau *conflation*. Proses *stemming* secara luas sudah digunakan di dalam kegiatan *Information retrieval* (pencarian informasi) untuk meningkatkan kualitas informasi yang didapatkan. Cara kerja *stemming* dapat dilakukan dengan merujuk ke kamus kata dasar maupun menggunakan aturan-aturan imbuhan (Indriyono, Utami, & Sunyoto, 2015).

Dalam penelitian ini untuk mendukung proses *preprocessing* digunakan algoritma *stemming* Nazief & Adriani. Algoritma ini mengubah setiap kata yang ditemukan menjadi kata dasar dan akan mengecek ketersediaannya dalam kamus.

3.2. Analisis Sentimen

Pada tahun 2001 sejarah analisis sentimen muncul pada jurnal karya dari Das, Chen, dan Tong. Bahasan yang mereka angkat yaitu menganalisis sentimen pasar. Analisis sentimen sendiri adalah cara untuk mengekstraksi pendapat, sentimen, evaluasi, dan emosi orang mengenai suatu topik tertentu yang tertulis dan menggunakan teknik pemrosesan bahasa alami. Adapun sejumlah karya-karya besar lainnya melakukan analisis sentimen yang fokus mengklasifikasikan mengenai sifat yang berlawanan yaitu

antara positif dan negatif. Dari pengertian tersebut, jadilah sebuah fakta yang menyebabkan beberapa penulis berpendapat bahwa istilah analisis sentimen mengacu pada tugas yang sempit atau spesifik. Namun saat ini banyak juga yang menafsirkan istilah analisis sentimen lebih luas, yang berarti cara pengkomputasian pendapat, sentimen, dan subjektivitas pada teks (Sentiaji & Bachtiar, 2014).

3.3. Metode Naïve Bayes Classifier

Naïve Bayes Classifier merupakan teknik prediksi berbasis probabilistik sederhana yang berdasarkan pada penerapan teorema atau aturan bayes, dengan asumsi independensi yang kuat pada fitur. Artinya bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama. *Naive Bayes Classifier* dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Naive Bayes Classifier* memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Naive Bayes Classifier* terbukti memiliki akurasi dan kecepatan yang tinggi pada saat diaplikasikan ke dalam database dengan data yang besar (Indriyono, Utami, & Sunyoto, 2015). Penjelasan mengenai metode Naïve Bayes Classifier adalah sebagai berikut (Wikanargo, 2016):

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \dots\dots\dots (1)$$

Persamaan (1) diatas yaitu $P(C_i|X)$ merupakan probabilitas C_i terjadi jika X sudah terjadi, $P(C_i)$

adalah probabilitas C_i dalam data dengan sifat *independent* terhadap X , X adalah kumpulan atribut, $P(X|C_i)$ adalah probabilitas X terjadi jika C_i sudah terjadi berdasarkan data training. Dengan $P(X)$ konstan maka didapat penyederhanaan sebagai berikut:

$$P(C_i|X) = P(X|C_i)P(C_i) \dots \dots \dots (2)$$

Persamaan (2) digunakan untuk mencari posisi probabilitas nilai dari data yang dicari berdasarkan data training. Model tersebut menunjukkan bahwa posisi probabilitas didapat dari nilai kemiripan data dibagi nilai data yang dicari dikalikan dengan nilai bukti dari data training.

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i) \times P(x_1|C_i) \times P(x_2|C_i) \times \dots \times P(x_n|C_i) \dots \dots \dots (3)$$

Persamaan model (3) digunakan untuk mendapatkan nilai probabilitas X , yang terjadi jika C_i sudah terjadi adalah dapat dilakukan dengan mengalikan nilai posisi probabilitas masing-masing kelas yang mendukung kelas tujuan. Nilai-nilai tersebut akan dihitung nilai posisi probabilitas data yang dicari masuk ke kelas tujuan.

Nilai yang didapatkan dari persamaan ini digunakan untuk mendapatkan nilai akhir C_i terjadi jika X sudah terjadi pada persamaan (2). Posisi probabilitas C_i terjadi jika X sudah terjadi pada model (2), yang terbesar adalah hasil yang diambil untuk menunjukkan kelasnya.

3.4. Jejaring Sosial Twitter

Twitter adalah sebuah situs web yang dioperasikan oleh Twitter Inc, Twitter merupakan jejaring sosial *mikroblogging* sehingga memungkinkan penggunanya untuk mengirim dan membaca pesan yang biasa disebut kicauan (*tweets*). *Tweets* merupakan teks tulisan yang terdiri atas 140 karakter dan ditampilkan pada halaman profil pengguna. *Tweets* bisa dilihat secara umum, namun pengirim yang memiliki *tweets* dapat membatasi pengiriman pesan ke daftar teman-teman mereka saja atau tertentu. Pengguna dapat juga melihat *tweets* dari penulis lain yang dikenal dengan pengikut (*follower*).

Semua pengguna dapat menerima dan mengirim kicauan melalui situs Twitter, telepon seluler, atau dengan pesan singkat (SMS) tergantung ketersediaannya di negara-negara tertentu. Situs ini pertama kali dibuat di San Bruno, California dekat San Francisco dan berbasis disana juga. Sedangkan kantor dan server Twitter berada di San Antonio, Texas dan Boston, Massachusetts.

Twitter dibentuk pada tahun 2006 oleh Jack Dorsey dan telah mendapatkan popularitas di seluruh dunia. Saat ini Twitter memiliki lebih dari 100 juta pengguna. Hal ini kadang-kadang digambarkan sebagai "SMS dari internet". Logo Twitter sendiri berupa seekor burung berwarna biru yang bernama "Larry the "Bird". Kelebihan Twitter dibandingkan dengan jejaring sosial lainnya yaitu (Supriyono, 2013) :

1. Twitter saat ini menjadi salah satu situs jejaring sosial yang paling sering dikunjungi di dunia setelah Google dan Facebook.
2. Pengguna Twitter yang terus mengalami pertumbuhan dan bahkan Indonesia tergolong dalam lima besar negara dengan pengguna Twitter terbesar.
3. Pembuatan akun Twitter gratis.
4. Hampir semua kalangan dapat menggunakan Twitter dan bisa mengoperasikannya.
5. Twitter membuat penggunaannya untuk selalu aktif dan bebas untuk melakukan *tweets*.
6. Teknologi Twitter tersedia dalam *mobile* yang membuat Twitter dapat dipantau secara fleksibel setiap waktu, kapan saja dan di mana saja.
7. Selain itu Twitter juga memungkinkan penggunaannya untuk menyambungkan akunnya dengan media sosial lain seperti *Facebook, Instagram, Soundcloud, Tumblr*, dan lain-lain.