

BAB 3

LANDASAN TEORI

Pada bab ini akan menjelaskan mengenai teori-teori yang digunakan dalam penyusunan penelitian ini. Teori-teori ini akan menjelaskan yang berhubungan dengan analisis sentimen masyarakat tentang *provider* Telekomunikasi Indonesia pada Twitter menggunakan *Naïve Bayes Classifier*.

5.1. Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung opini berpandangan atau beropini negatif atau positif (Rozi, Pramono, & Ac, Implementasi Opinion Mining (Analisis, 2012). Pengelompokan polaritas yang ada dalam dokumen, kalimat atau fitur aspek-aspek pendapat yang dikemukakan dalam dokumen, kalimat atau fitur entitas atau aspek ekspresi mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin akan berbeda makna dengan pernyataan yang sama pada *subject* yang berbeda (Lidya, Sitompul, & Efendi, 2015).

Analisis Sentimen dilakukan untuk melihat pendapat terhadap sebuah masalah, atau dapat juga digunakan untuk identifikasi hal yang sedang ramai dibicarakan. Analisis sentimen dalam penelitian ini adalah proses klasifikasi ke dalam dua kelas, yaitu kelas sentimen positif dan negatif. Besarnya pengaruh dan manfaat dari analisis sentimen, menyebabkan penelitian mengenai analisis sentimen berkembang pesat, bahkan di Amerika kurang lebih 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen. Manfaat analisis sentimen dalam dunia usaha antara lain untuk melakukan pemantauan terhadap sebuah produk. Analisis sentimen dapat digunakan sebagai alat bantu untuk melihat respon konsumen atau masyarakat terhadap suatu produk tertentu.

5.2. Jejaring Sosial Twitter

Twitter adalah sebuah media sosial yang digunakan sebagai alat untuk berbagi berbagai macam informasi mengenai banyak hal yang menjadi perhatian pengguna (Rivki & Bachtiar, 2017). *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. *Twitter* memberikan kemudahan akses terhadap akun, koneksi yang selalu terhubung dan perluasan jangkauan dengan *retweets* sehingga *twitter* memiliki potensi yang besar sebagai media *sharing*.

Twitter dibentuk pada tahun 2006 oleh Jack Dorsey dan telah mendapatkan perhatian dari seluruh penjuru dunia. Situs ini pertama kali dibuat di San Bruno, California dekat San Francisco. Sedangkan kantor dan Server berada di San Antonio, Texas dan Boston, Massachusetts. 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" (Rivki & Bachtiar, 2017).

Besarnya minat masyarakat Indonesia terhadap media sosial *twitter* dapat menjadi salah satu alasan mengapa banyak pula kalangan yang menggunakan media sosial tersebut untuk berbagai kepentingan. Bagi sebagian masyarakat *twitter* merupakan tempat untuk berkomunikasi dan berbagi informasi namun bagi beberapa pihak *twitter* dapat menjadi ajang promosi, pemberitaan, dan bisnis.

5.3. Naïve Bayes Classifier (NBC)

Naïve Bayes Classifier merupakan teknik pembelajaran mesin yang berbasis probabilistik, dengan metode sederhana tapi memiliki akurasi serta performansi yang tinggi dalam pengklasifikasian teks. Konsep dasar yang digunakan pada *Naïve Bayes Classifier* adalah *Teorema Bayes*. *Teorema keputusan Bayes* adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*). *Naïve Bayes* didasarkan pada asumsi

penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai *output*. Dengan kata lain, diberikan nilai *output*, probabilitas mengamati secara bersama adalah produk dari probabilitas individu (Ling, N, Kencana, & Bagu, 2014). Dasar teorema ini adalah (Han, Kamber, & Pei, 2012):

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

$P(H|X)$ = peluang H menghasilkan kondisi X

$P(X|H)$ = peluang X yang muncul saat kondisi H

$P(H)$ = peluang keseluruhan kemunculan H

$P(X)$ = peluang keseluruhan kemunculan X

Peluang H terjadi saat kondisi X ditentukan dari peluang kondisi X saat menghasilkan hasil H, dikali peluang H, dibagi peluang X. Pada pengaplikasiannya rumus ini berubah menjadi:

$$P(C_i|D) = \frac{P(D|C_i) \times P(C_i)}{P(D)} \quad (2)$$

$P(C_i|D)$ = peluang kelas C_i menghasilkan kondisi D

$P(D|C_i)$ = peluang kondisi D yang muncul di kelas C_i

$P(C_i)$ = peluang keseluruhan kemunculan kelas C_i

$P(D)$ = peluang keseluruhan kemunculan kondisi D

Pada rumus yang telah dijelaskan diatas, hasil dari peluang kelas C_i mengahsilkan kondisi D dapat dihitung dengan membagi hasil perkalian peluang D yang muncul di

kelas Ci dan peluang keseluruhan kemunculan kelas Ci dengan peluang keseluruhan kemunculan kondisi D.

5.4. Data Mining

Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang besar (Arga, Lestari, & Sutanta, 2017). *Data mining* akan memecahkan masalah dengan menganalisis data yang telah ada dalam basis data. *Data mining*, sering juga disebut *knowledge discovery in database (KDD)* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan pola keteraturan, pola hubungan dalam set data berukuran besar (Rohman, 2016). Dapat disimpulkan bahwa *data mining* merupakan proses pencarian pengetahuan berupa pola atau model data yang bermanfaat dari basis data yang besar dan kotor.

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dari dan bermanfaat bagi pemilik data (Musthafa, Suyono, & Sarosa, 2015). Hasil keluaran dari *data mining* ini dapat dijadikan untuk memperbaiki pengambilan keputusan di masa depan.

Ada beberapa teknik yang dimiliki *data mining* berdasarkan tugas yang bisa dilakukan (Dan, Sihwi, & Anggrainingsih, 2015), yaitu:

a) Deskripsi

Digunakan untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.

b) Estimasi

Model menggunakan *record* lengkap yang menyediakan nilai dari variabel *target* sebagai nilai prediksi.

c) Prediksi

Melakukan prediksi nilai dari hasil yang akan ada di masa mendatang.

d) Klasifikasi

Contoh kasusnya adalah penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

e) *Clustering*

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.

f) Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

Penelitian ini akan menggunakan fungsi klasifikasi dikarenakan sistem mengelompokan data didasarkan pada karakteristik alterntif. Dalam klasifikasi variabel,

tujuannya bersifat kategori. Istilah data mining dan *knowledge discovery in databases* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. KDD dapat dibagi menjadi beberapa tahap proses (Novianti & Aziz, 2015), yaitu:

a) Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsis\iten.

b) Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam suatu *database* baru.

c) Seleksi data (*data selection*)

Data yang ada pada *database* sering kali tidak semua dipakai, hanya data yang sesuai untuk dianalisis yang diambil dari *database*. Karena tidak semua tabel digunakan maka perlu dilakukan pembersihan data agar data yang akan diolah benar-benar relevan dengan yang dibutuhkan

d) Transformasi data (*data transformation*)

Data digabungkan kedalam format yang sesuai untuk diproses dalam *data mining*. Transformasi data merupakan proses perubahan atau penggabungan data ke dalam format yang sesuai.

e) Proses *data mining*

Proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dari data.

f) Evaluasi pola (*pattern evaluation*)

g) Presentasi pengetahuan (*knowledge presentation*).

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.