

BAB III

LANDASAN TEORI

3.1. *Data Mining*

Penambangan informasi atau data *mining* adalah upaya untuk mendapatkan informasi tersembunyi dari informasi yang sangat besar melalui perhitungan yang berbeda. Penambangan informasi bekerja sama dengan berbagai disiplin ilmu termasuk kerangka kerja kumpulan data, wawasan, dan contoh pengakuan. Penambangan informasi itu sendiri sangat penting untuk proses pengungkapan yang lebih besar, yang menggabungkan tugas pra-penanganan dan pasca-penanganan. Tugas pra-penanganan seperti ekstraksi informasi, papan informasi, pengurangan informasi dan pengembangan elemen. Penugasan pasca-penanganan seperti pemahaman contoh dan model, usia dan penegasan anggapan. Pengungkapan informasi dan penambangan informasi adalah dua siklus iteratif dan intuitif (Zaki dan Meira, 2014)[17]. Pengaturan prosedur penggalian informasi yang masuk akal untuk klasifikasi kasus data ke dalam kelas yang telah dikarakterisasi tergantung pada campuran spesifik dari karakteristik kejadian tersebut (Han et al., 2012)[18].

Data *mining* adalah penambangan data yang didefinisikan sebagai proses seleksi, eksplorasi, dan juga pemodelan dari suatu jumlah besar untuk membantu menemukan pola atau bentuk koordinasi yang biasanya tidak dilihat atau tidak diketahui dari pola tersebut. Menggunakan data yang mudah dimengerti dan bermanfaat untuk terbentuknya pengetahuan dari sejumlah data yang besar untuk mengetahui suatu pola [19]. Hasil dari pengolahan data *mining* memiliki tujuan untuk mengambil sebuah keputusan. Metode algoritma yang digunakan untuk memproses data *mining* adalah *Knowledge Discovery in Database* (KDD). Dalam proses data *mining* dibagi menjadi lima tahap yaitu:

1. Seleksi Data

Seleksi data adalah cara digunakan untuk memilih data dari banyak data yang telah dikumpulkan untuk menemukan informasi selanjutnya.

2. Pemilahan Data

Pemilahan data adalah cara yang dilakukan untuk memilah data secara berulang-ulang dilanjutkan dengan melakukan pemeriksaan data yang tidak konsisten dan akan dilakukan perbaikan data yang sekiranya salah dalam pengetikkan.

3. Pengelompokan Data

Pengumpulan informasi atau data dilakukan untuk mengelompokkan informasi yang ada sehingga informasi yang diambil untuk penjelasan sesuai dengan penggalian informasi.

4. Proses Pencarian Informasi

Proses pencarian informasi adalah proses klarifikasi data dengan metode yang benar dan terpilih untuk menemukan informasi penting dari data tersebut.

5. Evaluasi Data

Evaluasi data adalah proses pengolahan data yang terhapus dan memperoleh informasi untuk menentukan fakta yang sebenarnya dan menentukan apakah data tersebut bertentangan dengan fakta yang ada. Data yang dievaluasi berfungsi sebagai kesimpulan klarifikasi dari peristiwa yang sebenarnya.

3.2. *Text Mining*

Text Mining adalah metode penambangan data untuk mengeksplorasi kata-kata dengan bantuan *tools* program yang ada dalam komputer untuk menghasilkan suatu informasi terbaru. Mengambil data-data berbentuk teks sehingga dapat dianalisa keterhubungan antar dokumen satu dan lainnya [20]. Jadi penulis menggunakan teori *text mining* untuk membantu dalam proses pengkategorisasian teks dan pengelompokan teks sesuai dengan apa yang dicari pada penelitian ini. Penambangan teks dapat didefinisikan sebagai penemuan informasi baru dan sebelumnya tidak diketahui oleh komputer dengan secara otomatis mengekstraksi informasi dari berbagai sumber teks tidak terstruktur (Tan, 1999)[19].

Definisi singkat dari *Text Mining* adalah proses parsing teks untuk mengekstrak informasi yang berguna untuk tujuan tertentu. Perbedaan utama antara penambangan teks dan penambangan data terletak pada sumber data yang digunakan. Dalam data *mining*, data yang diekstraksi berasal dari skema terstruktur tertentu, sedangkan dalam *text mining* sumber data yang digunakan berasal dari teks yang relatif tidak terstruktur karena menggunakan tata bahasa manusia atau biasa disebut *natural language*. Secara umum, database dirancang untuk program dengan tujuan melakukan pemrosesan otomatis, sedangkan teks ditulis untuk dibaca langsung oleh manusia.

3.3. *Sentiment Analysis*

Analisis sentimen adalah proses menggali opini untuk meninjau segala sesuatu yang berhubungan dengan komputasi dengan hubungan kasus publik mengenai pendapat orang lain atau opini. Pendapat ini bersifat subjektif dan tidak cukup hanya mempertimbangkan pendapat satu atau dua orang saja untuk menghasilkan rangkuman hasil pendapat tersebut. Pendapat sentimen dibagi menjadi dua bidang, yaitu opini reguler dan opini komparatif. Opini reguler adalah opini yang hanya dilihat oleh satu entitas material, sedangkan opini komparatif adalah opini yang dilihat dari sudut pandang beberapa

entitas. Sehingga *sentiment analysis* menghasilkan berupa presentasi sentimen positif, negatif, atau netral. Maka dari itu penulis menggunakan teori *sentiment analysis* untuk membantu memahami untuk memperoleh opini sesuai dengan kategori yang dicari pada penelitian ini. Analisis sentimen adalah metode analisis untuk mengidentifikasi komentar pengguna. Menurut (GO, Huang, Bhayani, 2009)[21] analisis sentimen adalah bidang penelitian yang penting dan aktif yang didorong oleh pertumbuhan pesat media sosial dan peluang untuk mengakses wawasan berharga tentang berbagai masalah bisnis, global, dan sosial.

Analisis sentimen adalah pemeriksaan dengan bantuan komputer atas pendapat, perilaku, dan emosi seseorang terhadap suatu entitas. Dengan begitu banyak pendapat dari pengguna *Twitter* tentang suatu masalah, banyak penelitian analisis sentimen *Twitter* telah dilakukan. Pada kesempatan ini penulis akan memaparkan penelitian yang dilakukan.

3.4. *Preprocessing*

Preprocessing adalah proses pengolahan dari data mentah menjadi data bersih sehingga hanya akan memberikan informasi yang dibutuhkan saja [22]. Tujuan dari *preprocessing* adalah untuk memilih setiap kata dalam dokumen dan mengubahnya menjadi kata dasar yang memiliki arti terbatas dan proses ekstraksi teks akan memberikan hasil yang lebih memuaskan. Ini adalah langkah-langkah *preprocessing* yang akan dilakukan:

a. *Stopword*

Stopword adalah proses pembersihan data dengan menghilangkan kata-kata yang tidak dibutuhkan seperti kata yang tidak mempunyai arti. Contohnya *URL*, *hashtag* (#), dan *mentions* (@)

b. *Case Folding*

Case folding adalah suatu proses untuk mengubah huruf secara keseluruhan dari *uppercase* menjadi *lowercase*.

3.5. *Service quality*

Kualitas layanan atau *Service quality* adalah bentuk keterampilan yang terkait, tetapi tidak disamakan dengan kepuasan yang dihasilkan dari perbandingan harapan dan kinerja. Pada prinsipnya kualitas pelayanan dipandang sebagai syarat penting dan juga sebagai penentu daya saing dalam suatu pelayanan untuk membangun dan memelihara hubungan dengan pelanggan. *Service quality* atau kualitas pelayanan dapat dimanifestasikan dalam memenuhi kebutuhan dan harapan konsumen serta ketepatan penyampaiannya untuk mengimbangi ekspektasi konsumen. Lovelock dan Wirtz (2007) menjelaskan kualitas pelayanan suatu evaluasi kognitif jangka panjang dari konsumen terhadap pemberian layanan oleh perusahaan. Zeithaml, Bitner dan Gramler (2009) menemukan bahwa "kualitas layanan, persepsi komponen layanan suatu produk oleh pelanggan, juga merupakan penentu penting bagi kepuasan pelanggan." Hal ini menjadi penentu penting dalam mencapai kepuasan pelanggan.

Berdasarkan berbagai definisi tersebut dapat dipahami bahwa kualitas pelayanan dengan cara membandingkan persepsi para konsumen terhadap pelayanan yang secara nyata mereka peroleh atau terima dengan pelayanan yang sesungguhnya mereka inginkan terhadap atribut-atribut pelayanan suatu perusahaan [22]. Berdasarkan pengertian tersebut, maka disimpulkan bahwa *Service quality* bisa dilihat berdasarkan pelayanan yang telah dirasakan oleh konsumen, kemudian dibandingkan dengan apa yang konsumen inginkan. Sedangkan ketidakpuasan terjadi bila hasil tidak memenuhi keinginan. Kepuasan konsumen pada suatu produk atau jasa sebenarnya sesuatu yang tidak mudah untuk didapat jika suatu perusahaan tidak memahami dengan benar keinginan dari konsumennya. Untuk mengukur tingkat kepuasan

konsumen menggunakan lima dimensi dari *service quality*. Berikut lima dimensi dari *service quality*, yaitu:

1. *Tangible*

Tangible adalah kemampuan perusahaan dalam menampilkan bukti secara fisik kepada konsumen dengan baik. Contohnya tampilan fisik, fasilitas fisik, dan alat-alat pendukung yang digunakan dalam melakukan pelayanan.

2. *Reliability*

Reliability adalah kemampuan yang diberikan oleh karyawan untuk sebagai pelayanan kepada konsumen tanpa ada kesalahan dan ketepatan waktu sesuai waktu yang disepakati.

3. *Responsiveness*

Responsiveness adalah kemampuan karyawan dalam memberikan respon pelayanan dengan cepat kepada pelanggan untuk mengatasi masalah dan keluhan dengan memberikan solusi yang tepat.

4. *Assurance*

Assurance adalah jaminan yang diberikan oleh pengelola jasa kepada pelanggan meliputi jaminan keamanan, biaya, dan bersikap sopan santun.

5. *Empathy*

Empathy adalah kemampuan karyawan untuk merasakan secara emosional dengan memberi perhatian personal dan saling menghargai kepada pelanggan.

3.6. *Twitter*

Twitter adalah layanan di dalam jejaring sosial dan *microblog daring* yang bisa memberikan sebuah informasi, pengguna *twitter* bisa mengirim dan membaca pesan dalam bentuk teks [23]. *Twitter* adalah media sosial yang

memungkinkan penggunaannya untuk mengirim dan membaca pesan *tweet* dalam bentuk teks, gambar, atau video. Jejaring sosial *Twitter* berbeda dengan jejaring sosial lainnya, terutama dalam hal menulis status atau *tweet*. Di media sosial selain *Twitter*, tidak ada batasan 6 karakter yang dapat diketik, sedangkan *Twitter* hanya memiliki 280 karakter yang dapat diketik sebagai status atau *tweet*. *Twitter* bersifat publik, artinya segala sesuatu yang ditulis atau dibagikan dapat dilihat oleh semua pengguna lain, tetapi pengguna *Twitter* dapat membatasi pengiriman *tweet* yang hanya dilihat oleh temannya atau yang biasa dikenal sebagai pengikut.

Twitter memiliki fungsi utama yaitu Anda dapat menulis status atau *tweet* dan mengirim pesan ke pengguna lain. Berikut fitur-fitur yang ada pada *twitter*:

a. *Home*

Pada *home twitter*, pengguna *twitter* dapat melihat *tweet* dari orang yang telah diikuti oleh pengguna.

b. *User Profile*

Pada profil pengguna terdapat identitas pengguna yang bisa dilihat oleh pengguna lain dengan untuk menarik perhatian agar saling mengikuti.

c. *Followers*

Followers adalah istilah dari pengikut dari akun seseorang kepada pengguna lain. Ketika pengguna membagikan *tweet*, maka akan terlihat di bagian *home followers*.

d. *Following*

Following adalah istilah dari seorang pengguna yang mengikuti akun *twitter* seseorang. Ketika akun yang di ikuti membagikan *tweet*, maka akan terlihat pada *home* pengguna.

e. *Direct Message*

Direct message seperti layaknya fitur *chatting* yang berfungsi untuk mengirimkan pesan dari pengguna ke pengguna lainnya dan bersifat *private message* antara kedua pengguna.

f. *Hashtag*

Simbol dari *hashtag* adalah “#” yaitu untuk mempermudah pengguna mencari suatu *tweet* sejenis yang pengguna lain bagikan.

g. *Mentions*

Simbol dari *mentions* adalah “@” yang berfungsi untuk melakukan percakapan antara seorang pengguna dengan banyak pengguna lainnya. Tidak bersifat *private message* antara kedua pengguna. Jadi penggunaan *mentions* dalam percakapan bisa dilihat oleh banyak pengguna lainnya.

h. *Trending Topic*

Trending topic adalah fitur yang memperlihatkan kejadian yang sedang banyak dibahas (viral) oleh para pengguna *twitter*. Para pengguna menggunakan *hashtag* untuk mengirimkan *tweet*. Jika respon pengguna lain ikut serta dalam membahas *tweet* dari *hashtag* yang sama dan memperoleh *hashtag* terbanyak, maka akan masuk *trending topic*.

3.7. Metode Klasifikasi *Sentistrength*

Metode klasifikasi *Sentistrength* adalah metode pemberian bobot atau nilai pada emosi dan membagi kelas emosi menjadi kelas positif, negatif, dan netral sesuai dengan nilai yang terdapat dalam surat yang disampaikan. [10]. Metode tersebut dapat memperkirakan nilai sentimen dari beberapa kamus-kamus yang sudah diberi bobot masing-masing. Bobot tersebut diberi nilai dari -5 (memiliki nilai sentimen negatif yang sangat kuat) sampai dengan 5 (memiliki nilai sentimen positif yang sangat kuat). Berikut kamus - kamus

yang digunakan sebagai patokan untuk menentukan bobot sentimen antara lain:

1. Kamus Sentimen

Pada kamus sentimen ini sudah diberi bobot pada kata - kata yang memiliki sentimen positif, negatif, atau netral. Contohnya sebagai berikut:

Kata	Bobot
Sejahtera	5
Baik	4
Seimbang	5
Aduh	-2
Hancur	-5

2. Kamus Kiasan

Pada kamus kiasan adalah suatu kata ungkapan yang memiliki makna tidak sama dari kata aslinya. Contohnya sebagai berikut:

Kata	Bobot
Keras kepala	-4
Tulang punggung	2
Bajing loncat	-4
Panjang tangan	-5
Turun tangan	3

3. *Boosterwords*

Boosterwords adalah kata-kata yang dapat menambah atau mengurangi nilai makna emosional dari kata-kata yang berdekatan. Contohnya sebagai berikut:

Kata	Bobot
Sangat	2
Mungkin	-1
Mending	1
Kurang	-2
Lebih	1

4. Kata Negasi

Negasi suatu kata adalah kata yang dapat mengubah arti kata tersebut. Misalnya, kata "baik" dengan nilai emosional positif didahului dengan kata negatif. Misalnya, "tidak baik". Kata tersebut menjadi sebuah emosi yang bernilai negatif. Contoh kata negatif seperti:

No	Kata
1	Tidak
2	Kurang
3	Belum
4	Bukan
5	Pantang

5. Kata Tanya

Kata tanya dapat mengubah suatu makna nilai opini seseorang dalam menyampaikan tulisan di *twitter*. Contohnya “kenapa barangnya rusak?”. Walaupun pada kalimat tersebut terdapat kata “rusak”, tetapi kalimat tersebut tidak termasuk ke dalam positif atau negatif, melainkan masuk ke dalam kategori netral. Contoh beberapa kata tanya sebagai berikut:

No	Kata
1	Apakah
2	Bagaimana
3	Berapa
4	Siapa
5	Kenapa

3.8. *Corpus*

Corpus didefinisikan sebagai sekumpulan teks tertulis dan lisan dalam format data yang dapat dibaca oleh banyak mesin dan ditulis dalam berbagai bentuk informasi linguistik. *Corpus* adalah kumpulan teks atau kata-kata yang membentuk kamus. *Corpus* juga dapat dipahami sebagai kumpulan teks dalam kumpulan teks linguistik elektronik yang dipilih menurut kriteria eksternal untuk mewakili sejumlah besar bahasa. *Corpus* adalah kamus yang dapat digunakan karena dikumpulkan dari berbagai sumber, termasuk materi cetak dan media elektronik, termasuk frasa, clausa, dan wacana. [24].