

## BAB 2

### TINJAUAN PUSTAKA

#### 2.1 Analisis Sentimen

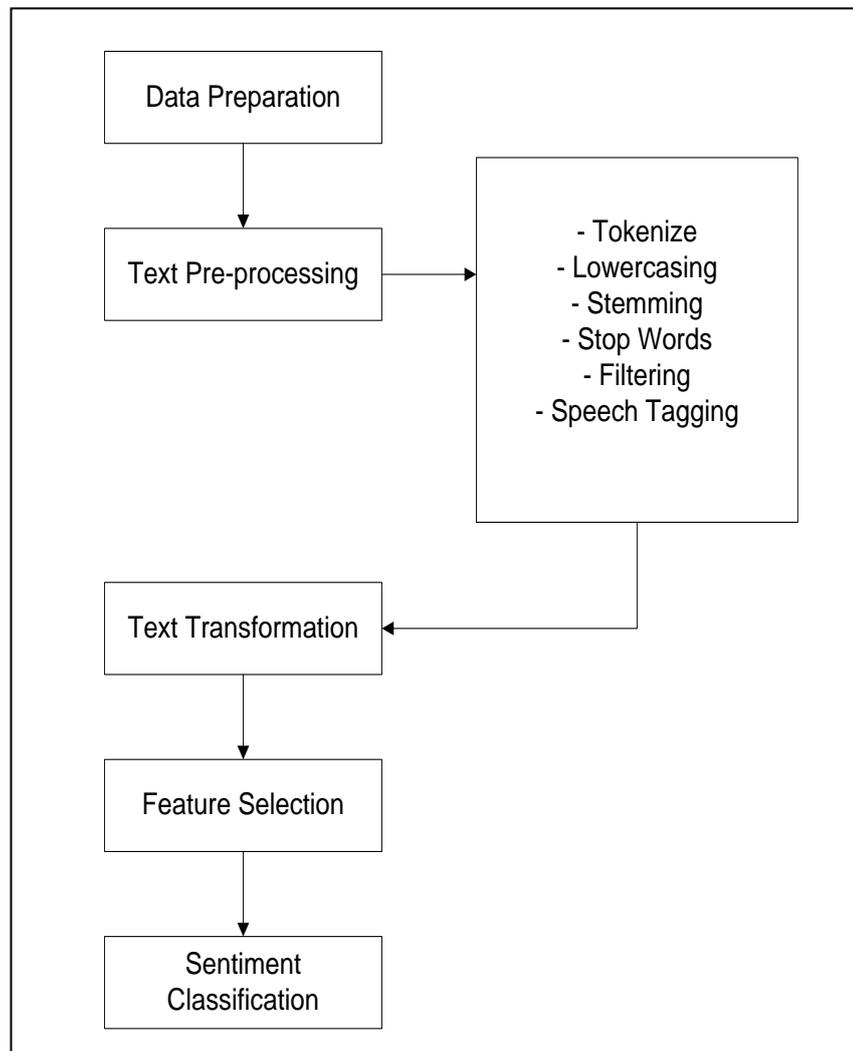
Analisis sentimen adalah proses penghitungan dan pengkategorian opini secara komputatif yang diungkapkan dalam sebuah kalimat, terutama untuk menentukan apakah sikap seseorang terhadap suatu topik, produk, dll., adalah positif, negatif, atau netral. Ada banyak istilah atau nama yang sedikit berbeda, misalnya, *opinion mining*, *opinion extraction*, *sentimen mining*, *subjectivity analysis*, *affect analysis*, *emotion analysis*, *review mining*, dan lain-lain. Dari beragam istilah tersebut, kalangan akademisi dan industri lebih sering menggunakan istilah analisis sentimen. Riset mengenai analisis sentimen mulai marak sejak paper dari B. Pang dan L. Lee [1] keluar. Secara umum, analisis sentimen ini dibagi menjadi 2 kategori besar, yaitu :

1. *Coarse-grained sentiment analysis*
2. *Fined-grained sentiment analysis*

*Coarse-grained sentiment analysis* adalah proses analisis pada level dokumen. Singkatnya adalah kita mencoba mengklasifikasikan orientasi sebuah dokumen secara keseluruhan. Orientasi ini ada 3 jenis: Positif, Netral, Negatif. Akan tetapi, ada juga yang menjadikan nilai orientasi ini bersifat kontinu / tidak diskrit. *Sementara fined-grained sentiment analysis* adalah proses analisis pada level sebuah kalimat pada suatu dokumen

Dalam analisis sentimen dasar, banyak penelitian menggunakan teknik *machine learning* seperti mesin *Naïve Bayes*, *Maximum Entropy*, dan *Support Vector* karena algoritma ini cenderung mengungguli algoritma lain dalam konteks klasifikasi teks. Selain itu, ada juga teknik *lexicon-based* yang menggunakan kamus *lexicon* untuk melakukan penilaian terhadap kata. Pada kamus *lexicon*, kata yang akan dianalisis dari corpus telah ditentukan dan dipasangkan dengan nilai polaritasnya. Ada juga teknik *hybrid* yang merupakan gabungan dari dua teknik pendekatan sebelumnya.

Adapun langkah-langkah melakukan analisis sentimen secara garis besar dapat dilihat dalam gambar berikut.

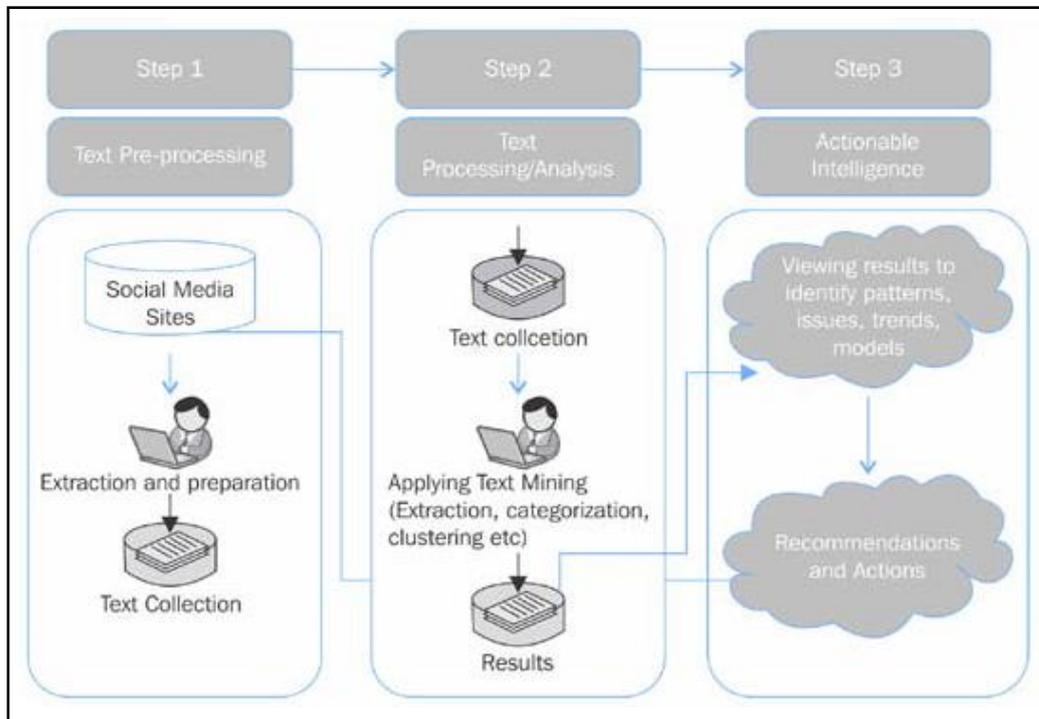


*Gambar 2.1 Gambaran Umum Proses Analisis Sentimen*

Penjelasan dari gambar 2.1 yaitu, langkah pertama yang harus dilakukan adalah mempersiapkan data, setelah itu dilakukan penormalan data tersebut dengan tujuan menghilangkan kata-kata yang tidak diperlukan dan mempermudah perhitungan sentimen. Setelah teks diproses menjadi normal, lalu dipilih sebuah kata yang dianggap sebagai sentimen teks tersebut dan di klasifikasikan apakah sentimen tersebut positif, negatif atau netral.

### 2.1.1 Analisis Sentimen Twitter Menggunakan R

Penelitian ini merupakan lanjutan dari proses analisis sentimen yang digunakan oleh Sharan Kumar [9] menggunakan R Studio dengan *dataset* yang diambil dari twitter.



Gambar 2.2 Analisis Sentimen Twitter Menggunakan R

Langkah tersebut dimulai dari menghubungkan *R Studio* dengan *Twitter* menggunakan *Twitter API* lalu mengumpulkan *tweet* dengan cara *crawling*. Tahap selanjutnya berurutan seperti langkah-langkah yang ditunjukkan oleh gambar 2-1. Setiap kata yang sudah dibersihkan melalui proses *pre-processing* akan diberikan label berupa angka dengan jarak parameter angka dari -2 sampai 2. Angka -2 berarti kata tersebut memiliki sentimen ekstra negatif. Angka -1 berarti kata tersebut memiliki sentimen negatif. Angka 0 berarti kata tersebut mengandung sentimen negatif atau positif. Angka 1 berarti kata tersebut memiliki sentimen positif. Angka 2 berarti kata tersebut memiliki sentimen ekstra positif. Setelah keseluruhan kalimat dianalisis maka dijumlahkanlah total nilai yang didapat pada setiap *tweet*nya.

## 2.2 Pendeteksian Kalimat Ironi

Selama ini orang-orang sering salah mengartikan ironi sebagai sarkasme. Sarkasme adalah penggunaan kata-kata pedas untuk menyakiti hati orang lain seperti cemoohan atau ejekan kasar. Sementara ironi adalah bentuk wicara yang pintar di mana pembicara atau penulis mengatakan atau menulis kebalikan dari apa yang mereka maksud [2]. Ketika berbicara, orang-orang sering menggunakan penekanan nada yang keras dan petunjuk gestural tertentu seperti memutar mata, gerakan tangan untuk mengungkapkan ironi. Dalam data tekstual, penekanan nada dan petunjuk gestural yang hilang, membuat pendeteksian ironi sangat sulit bahkan bagi manusia sekalipun. Analisis sentimen pendeteksian ironi dalam wilayah *Natural Language Processing* (NLP) berkembang sangat pesat dengan penelitian mulai dari tingkat kata, tingkat klasifikasi frase dan kalimat, tingkat dokumen, dan tingkat klasifikasi konsep. Penelitian analisis sentimen berkembang mencari cara yang lebih efisien dan akurasi yang lebih baik untuk pendeteksian ironi dengan pada sebuah data tekstual. Pendeteksian Sentimen ironi diklasifikasikan ke dalam tiga kategori berdasarkan fitur teks yang digunakan untuk klasifikasi yaitu, bersifat leksikal, pragmatis, dan hiperbolik.

Lunando dan Purwarianti [4] menyatakan bahwa hampir semua kalimat ironi berbentuk kalimat positif yang memiliki makna negatif, walau bukan tidak mungkin terdapat kalimat ironi bermakna positif. Kalimat ironi pada topik makanan, kehidupan, dan kesehatan jarang sekali ditemukan, hanya terdapat 2 dari 100 kalimat. Akan tetapi pada topik pemerintahan, merek, atau politik, kalimat ironi lebih sering ditemukan. Dari 100 kalimat, terdapat 18 kalimat ironi. Penelitian tentang pendeteksian sentimen ironi dalam konteks analisis sentimen telah banyak dilakukan. Diantaranya, Riloff [3] mendeteksi ironi jika memperlihatkan perbedaan yang jelas antara sentimen positif terhadap situasi yang negatif. Seperti "*Hebat sekali pelayanan toko itu, membuat konsumen menunggu begitu lama.*" Lunando dan Purwarianti [4] mengusulkan metode unigram+negativity+interjection dimana kalimat bersifat ironi tersebut biasanya diawali kata-kata yang mengandung interjection words seperti "aha", "bah", "nah", "wew", "wow", "yay", "uh", dll. Sementara itu untuk mendeteksi kalimat ironi dibutuhkan

informasi lain seperti tanggapan orang lain terhadap sebuah topik, dan sejarah sentimen seseorang di masa lalu.

Sentimen yang seseorang ungkapkan di masa lalu dan pendapat orang lain terhadap sebuah topik perlu diklasifikasikan. Sehingga ketika seseorang memberikan sentimen positif diantara mayoritas sentimen negatif membantu memberikan informasi kepada kita untuk memahami jika orang tersebut mempunyai kecenderungan sedang ironi. Anupam [5] mengusulkan metode *Historical Tweet-based Predictor* yang mengidentifikasi ironi jika sentimen yang diekspresikan terhadap entitas dalam target tweet setuju dengan sentimen yang diekspresikan oleh penulis terhadap entitas itu dimasa lalu.



Gambar 2.3 Contoh Historical Tweet-based Predictor

Penelitian [5] tergantung kepada entitas target sentimen orang tersebut, sehingga sulit mendefinisikan sebuah ujaran bersifat ironi atau tidak apabila tidak ada entitas target yang sama pada sejarah ujaran orang tersebut.

### 2.3 Logika Fuzzy

Logika *fuzzy* adalah perpanjangan dari logika *Boolean* oleh Lotfi Zadeh pada tahun 1965 berdasarkan pada teori matematika dari himpunan *fuzzy*, yang merupakan generalisasi teori himpunan klasik. Dengan memperkenalkan pengertian derajat dalam verifikasi suatu kondisi, sehingga memungkinkan suatu kondisi menjadi dalam keadaan selain benar atau salah, logika *fuzzy* memberikan nilai yang sangat berharga fleksibilitas untuk penalaran, yang memungkinkan untuk memperhitungkan ketidakakuratan dan ketidakpastian. Logika *fuzzy* telah digunakan dalam bidang manajemen transportasi, analisis sentimen, pencarian informasi, dan ekstraksi informasi. Logika *fuzzy* adalah modifikasi yang tepat dari konseptualisasi bersama dalam kategori tertentu, yang merupakan format yang dimengerti manusia dan dapat dibaca oleh mesin. Untuk spesifikasi lebih lanjut dari proses analisis sentimen, logika *fuzzy* dapat diperkenalkan. Oleh karena itu, analisis sentimen dengan bantuan logika *fuzzy* yang berurusan dengan penalaran dan memberikan pandangan lebih dekat ke nilai sentimen yang tepat akan membantu pengklasifikasian sejarah ujaran seseorang.

Logika *fuzzy* membantu penelitian ini untuk memberi nilai kecenderungan sentimen seseorang berdasarkan sejarah ujarannya. Algoritma *fuzzy* mampu membuat kriteria berdasarkan himpunan data yang batasnya tidak tegas. Contohnya adalah ketika mayoritas ujaran seseorang nilai kecenderungan positif dan negatifnya tidak jelas yang disebabkan oleh banyaknya ujaran yang tidak bisa dinilai oleh teknik analisis sentimen atau bersifat netral. Sementara jumlah ujaran orang tersebut terus bertambah, sehingga batasan yang dibuat apabila kita menggunakan teknik perhitungan matematika sederhana hasilnya kurang maksimal.

### 2.3.1 *Fuzzy Mamdani*

Di dalam perhitungan logika *fuzzy* terdapat beberapa metode, yaitu metode Tsukamoto, metode Mamdani, dan metode Sugeno. Setiap metode tersebut memiliki hasil perhitungan yang berbeda. Dalam kasus ini, masalah yang timbul adalah bagaimana cara menerapkan metode *fuzzy* Mamdani untuk menentukan sebuah ujaran bermakna ironi atau bukan atas kesejarahan seseorang. Metode *fuzzy* mamdani diperkenalkan oleh Ebrahim Mamdani pada tahun 1975. Metode *Fuzzy* Mamdani merupakan salah satu bagian dari *Fuzzy Inference System* yang berguna untuk penarikan kesimpulan atau suatu keputusan terbaik dalam permasalahan yang tidak pasti [10]. Kelebihan pada *Metode Fuzzy Mamdani* adalah lebih spesifik, artinya dalam prosesnya Metode *Fuzzy* Mamdani lebih memperhatikan kondisi yang akan terjadi untuk setiap daerah *fuzzynya*, sehingga menghasilkan hasil keputusan yang lebih akurat [11]. Karena jumlah ujaran seseorang terus bertambah dan tidak ada ukuran yang pasti untuk menentukan bahwa seseorang tergolong positif atau negatif berdasarkan sejarah ujarannya, maka dari itu metode *fuzzy* mamdani dianggap tepat digunakan untuk penelitian ini. Untuk mendapatkan output dari proses *fuzzy* mamdani ini, diperlukan 4 tahapan yaitu: Pembentukan himpunan *fuzzy*, Aplikasi fungsi implikasi (aturan), Komposisi aturan, Penegasan (*defuzzy*).

### 2.3.2 *Himpunan Fuzzy*

Pada himpunan tegas di setiap elemen dalam semestanya akan selalu ditentukan secara tegas apakah elemen itu betul merupakan anggota himpunan tersebut atau bukan. Akan tetapi dalam kenyataannya hampir tidak semua himpunan akan terdefinisi secara tegas [11]. Misalnya saja himpunan mahasiswa cerdas, dalam contoh ini tidak bisa dinyatakan dengan tegas karena kita lihat tidak ada yang dijadikan ukuran pasti untuk tingkat kecerdasan dari seseorang. Oleh karena itu perlu didefinisikan suatu himpunan *fuzzy* yang bisa menyatakan kejadian tersebut. Himpunan *fuzzy* menurut Kusumadewi [12] memiliki 2 atribut, yaitu:

1. Linguistik, yaitu penamaan suatu kelompok yang mewakili suatu keadaan atau kondisi tertentu dengan menggunakan bahasa alami, seperti: JAUH, SEDANG, DEKAT.
2. Numeris, yaitu suatu nilai (angka) yang menunjukkan ukuran dari suatu variabel, seperti: 10, 40, 80, dan sebagainya.

Hal – hal yang perlu diketahui dalam memahami sistem *fuzzy*, yaitu:

a) Variabel *Fuzzy*

Variabel *fuzzy* merupakan variabel yang akan dibahas dalam suatu sistem *fuzzy*, seperti: tingkat kecenderungan, potensi ironi, dan sebagainya.

b) Himpunan *Fuzzy*

Himpunan *fuzzy* merupakan suatu kelompok yang mewakili suatu keadaan tertentu dalam suatu variabel *fuzzy*.

c) Semesta pembicaraan

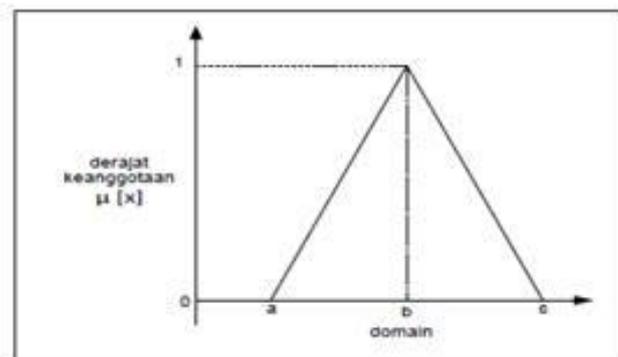
Semesta pembicaraan adalah keseluruhan nilai yang diperbolehkan untuk dioperasikan dalam suatu variabel *fuzzy*. Nilai semesta pembicaraan dapat berupa bilangan positif maupun negatif.

d) Domain

Domain himpunan *fuzzy* adalah keseluruhan nilai yang diperbolehkan untuk dioperasikan dalam suatu himpunan *fuzzy*. Nilai domain dapat berupa bilangan positif maupun negatif.

### 2.3.3 Fungsi Keanggotaan

Fungsi keanggotaan adalah suatu kurva yang menunjukkan pemetaan titik – titik input data ke dalam nilai keanggotaannya yang memiliki interval antara 0 sampai 1. Ada beberapa representasi kurva yang terdapat dalam fungsi keanggotaan *fuzzy* seperti linear, segitiga, trapesium, dll. Pada penelitian ini representasi kurva yang digunakan adalah berbentuk segitiga karena jenis representasi kurva tersebut sangat cocok diterapkan untuk kondisi 3 fungsi keanggotaan seperti kasus dalam penelitian ini terdapat 3 variabel masukan dengan 3 fungsi keanggotaan untuk membentuk variabel keluaran berupa tingkat kecenderungan sejarah ujaran seseorang yaitu masukan ujaran negatif, masukan ujaran netral, dan masukan ujaran positif. Dan ketiga variabel masukan tersebut memiliki 3 fungsi keanggotaan yang sama yaitu rendah, sedang, dan tinggi.



Gambar 2.4 Representasi Kurva Segitiga

Fungsi Keanggotaan :

$$\mu[x] = \begin{cases} 0; & x \leq a \text{ atau } x \geq c \\ \frac{(x-a)}{(b-a)}; & a \leq x \leq b \\ \frac{(c-x)}{(c-b)}; & b \leq x \leq c \end{cases}$$

Keterangan:

a = nilai domain terkecil yang mempunyai derajat keanggotaan nol

b = nilai domain yang mempunyai derajat keanggotaan satu

c = nilai domain terbesar yang mempunyai derajat keanggotaan nol

x = nilai input yang akan di ubah ke dalam bilangan *fuzzy*

#### 2.3.4 Fungsi Implikasi

Setiap aturan (proposisi) pada basis pengetahuan *fuzzy* akan berhubungan dengan suatu relasi *fuzzy*. Dimana bentuk umum dari aturan yang digunakan dalam fungsi implikasi adalah sebagai berikut:

IF x is A THEN y is B

Dengan x dan y adalah skalar, dan A dan B adalah himpunan *fuzzy*. Proposisi yang mengikuti IF disebut sebagai anteseden, sedangkan proporsi yang mengikuti THEN disebut sebagai konsekuen.

#### 2.3.5 Komposisi Aturan

Apabila sistem terdiri dari beberapa aturan, maka inferensi diperoleh dari gabungan antar aturan. Ada tiga metode yang digunakan dalam melakukan inferensi sistem *fuzzy*, yaitu: *max*, *additive* dan probabilistik OR (probor).

##### a. Metode *Max* (*Maximum*)

Pada metode ini, solusi himpunan *fuzzy* diperoleh dengan cara mengambil nilai maksimum aturan, kemudian menggunakannya untuk memodifikasi daerah *fuzzy*, dan mengaplikasikannya ke output dengan menggunakan operator OR (union). Jika semua proposisi telah dievaluasi, maka output akan berisi suatu himpunan *fuzzy* yang merefleksikan kontribusi dari tiap-tiap proposisi.

b. Metode *Additive (Sum)*

Pada metode ini, solusi himpunan *fuzzy* diperoleh dengan cara melakukan bounded-sum terhadap semua output daerah *fuzzy*.

c. Metode Probabilistik OR (probor)

Pada metode ini, solusi himpunan *fuzzy* diperoleh dengan cara melakukan product terhadap semua output daerah *fuzzy*.

### 2.3.6 Penegasan (defuzzyfikasi)

Pengendali logika *fuzzy* harus mengubah variabel keluaran *fuzzy* menjadi nilai-nilai tegas yang dapat digunakan untuk mengendalikan sistem. Proses ini disebut penegasan (*defuzzification*). Sehingga jika diberikan suatu himpunan *fuzzy* dalam range tertentu, maka harus dapat diambil suatu nilai crisp tertentu sebagai output. Terdapat beberapa jenis metode pada defuzzifikasi pada komposisi aturan Mamdani seperti metode centroid, Metode Bisektor, Metode Mean of Maksimum (MOM), Metode Largest of Maximum (LOM), Metode Largest of Maximum (LOM), Metode Smallest of Maximum (SOM). Pada penelitian ini defuzzyfikasi yang digunakan adalah metode centroid karena hasil nilai dari penggunaan metode defuzzyfikasi diambil dari titik pusat pada daerah *fuzzy*.