



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Media Sosial**

Media sosial adalah forum baru yang membawa orang untuk bertukar ide, terhubung dengan menggunakan internet, dan memobilisasi untuk suatu tujuan dan mencari nasihat (Amedie, 2017). Salah satu media sosial yang populer pada zaman sekarang dan dipakai dalam penelitian ini adalah *Instagram*.

*Instagram* adalah bentuk komunikasi yang relatif baru di mana pengguna dapat dengan mudah membagikan pembaruan mereka dengan mengambil foto dan mengubah foto tersebut dengan menggunakan filter (Hu, Manikonda, & Kambhampati, 2014). Di aplikasi *Instagram* juga terdapat *caption* yang berisi teks untuk menjelaskan foto atau video yang sudah dibagikan oleh pengguna.

#### **2.2 Data Mining**

Seiring dengan perkembangan teknologi yang semakin pesat, *database* sebuah perusahaan maupun organisasi pun semakin besar. Data yang semakin besar dan banyak secara volume ini harus dimanfaatkan dengan baik oleh perusahaan maupun organisasi agar dapat memberikan keunggulan dibanding perusahaan maupun organisasi yang lain. Proses untuk mengolah data dengan volume yang besar membutuhkan cara-cara dan teknologi yang canggih agar pengolahan data tersebut dapat menghasilkan informasi yang bermakna dan bernilai.

Metode-metode pengolahan data pun semakin berkembang. Salah satu metode yang baru berkembang pada sebuah perusahaan maupun organisasi adalah *data mining*.

*Data mining* adalah sebagai proses yang mencakup penggalian informasi menarik, dapat ditafsirkan, dan bermanfaat dari data mentah (Madni, Anwar, & Shah, 2017).

### **2.3 Text Mining**

Banyaknya jumlah teks yang ada melalui internet dalam bentuk perpustakaan digital, *repository*, dan informasi tekstual lainnya seperti *blog*, *email* menyebabkan tantangan dalam menentukan pola dan bagaimana untuk cara mengolah data berbentuk teks dengan volume yang sangat besar. Alat *data mining* tradisional tidak mampu dalam mengolah tekstual data yang sangat besar karena memerlukan waktu dan upaya dalam mengekstrak informasi.

Oleh karena itu, ada metode yang dinamakan *text mining*. *Text mining* adalah proses untuk mengekstrak pola yang menarik dan signifikan untuk mendapatkan informasi dari data tekstual (Talib, Kashif, Ayesha, & Fatima, 2016).

### **2.4 Analisis Sentimen**

Dalam penggunaan metode *text mining*, kita dapat mengetahui sebuah informasi mengenai sentimen pembicara dalam sebuah teks yang ia kirim melalui media sosial maupun *platform* yang lain.

Analisis sentimen adalah klasifikasi teks yang mengklasifikasikan dokumen berdasarkan arah objek. Dengan demikian, bagian terpenting dalam *text mining* adalah pemrosesan bahasa. Analisis sentimen juga dikenal sebagai proses mendeteksi polaritas kontekstual dari teks dan menentukan apakah teks ini bersifat positif, negatif, atau netral. Hal tersebut digunakan untuk memperoleh penilaian, emosi, atau sikap pembicara (Faiz & Dahab, 2017).

## **2.5 Text Preprocessing**

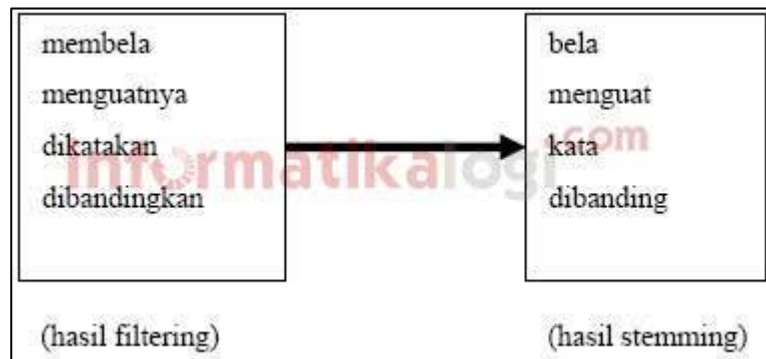
Tahap *pre-processing* atau praproses data merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain (Mujilahwati, 2016). Tahap *preprocessing* dibagi menjadi 4 tahap, yaitu:

### 1. *Case Folding*

Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu, peran *Case Folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar yaitu mengubah semua huruf dalam dokumen menjadi huruf kecil.

### 2. *Stemming*

Tahap *stemming* adalah tahap dimana melakukan pengelompokan kata-kata yang lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau form yang berbeda karena mendapatkan imbuhan yang berbeda. Berikut adalah contoh dari *stemming*:

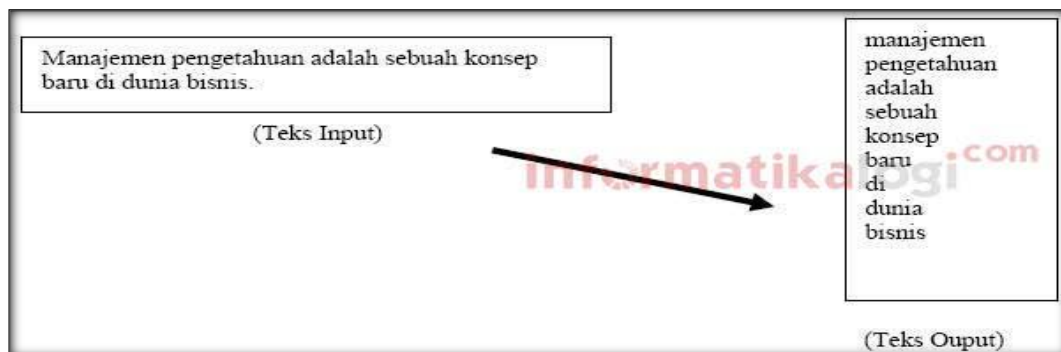


**Gambar 2.1 Contoh Stemming**

Sumber : (Informatikalogi, 2017)

### 3. *Tokenizing*

Tahap *tokenizing* adalah tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Sebagai contoh dapat dilihat dari gambar sebagai berikut:



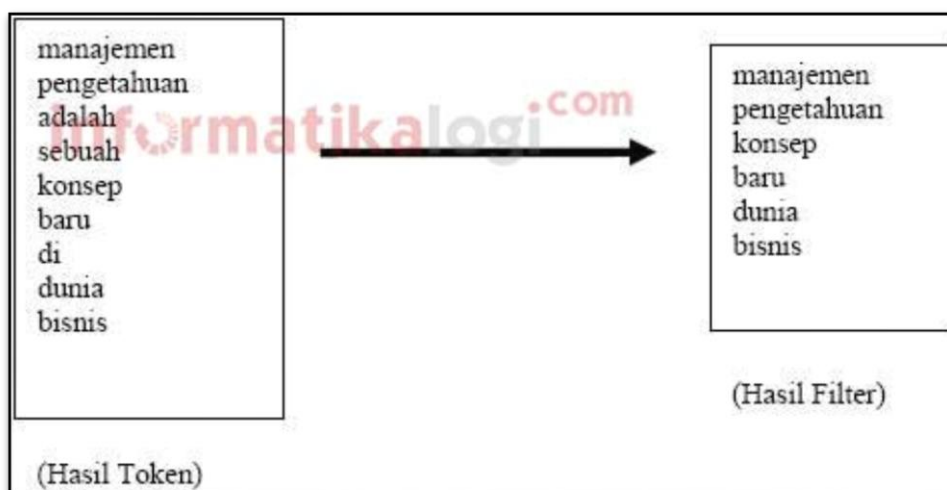
**Gambar 2.2 Contoh Tokenizing**

Sumber : (Informatikalogi, 2017)

### 4. *Filtering*

Tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil token. Pada tahap ini, dapat menggunakan dua algoritma, yaitu *stoplist* dan *wordlist*. *Stoplist* adalah membuang kata yang kurang penting

sedangkan *wordlist* adalah menyimpan kata penting. Contoh *stopwords* adalah “yang”,”dan”,”di”,”dari” dan lain-lain (Rahutomo & Ririd, 2019). Berikut adalah contoh dari *filtering*:



**Gambar 2.3 Contoh Filtering**

Sumber : (Informatikalogi, 2017)

## 2.6 Naïve Bayes Classifier

Klasifikasi *Naïve Bayes* adalah salah satu teknik klasifikasi dimana teknik tersebut mengklasifikasikan sebuah teks/kalimat ke dalam suatu kelas tertentu. Klasifikasi *Naïve Bayes* adalah algoritma probabilistik yang menghitung probabilitas setiap kata dalam teks/kalimat dan kata dengan probabilitas tertinggi dianggap sebagai output (Suppala & Rao, 2019).

Secara umum, teorema Bayes dapat dikatakan titik data  $x$  adalah kelas yang belum diketahui dan kelas  $c$  merupakan suatu kelas spesifik. Rumus Persamaan Teroema Bayes dapat dilihat pada rumus 2.1 berikut:

$$P(C|X) = \frac{P(x|c)P(c)}{P(x)}$$

### Rumus 2.1 Persamaan Teorema Bayes

$P(C|X)$  merupakan probabilitas hipotesis berdasarkan kondisi.  $P(x|c)$  merupakan probabilitas berdasarkan kondisi pada hipotesis.  $P(c)$  merupakan probabilitas hiperior.  $P(x)$  merupakan probabilitas  $c$ .

Contoh mudah menggunakan *Naïve Bayes* dapat dilihat pada tabel 2.1 berikut (Dey, Chakraborty, Biswas, Bose, & Tiwari, 2016):

**Tabel 2.1 Contoh Algoritma *Naïve Bayes***

<i>Set</i>	<i>Document</i>	<i>Review Sentence</i>	<i>Class</i>
<i>Training Set</i>	1	Saya suka film itu	Positif
	2	Film itu sangat bagus. Cerita yang indah.	Positif
	3	Akingnya jelek tapi heronya bagus. Keseluruhan filmnya bagus.	Positif
	4	Lagu yang indah. Tetapi ending yang membosankan	Negatif
<i>Test Set</i>		Saya suka filmnya. Tetapi tempat syuting yang membosankan. Keseluruhan filmnya bagus.	Positif

### 2.7 Support Vector Machine

*Support Vector Machine* adalah algoritma canggih dengan fondasi teoritis yang kuat berdasarkan teori *Vapnik-Chervonenkis*. *Support Vector Machine* memiliki regulasi yang kuat, regulasi tersebut mengacu pada generalisasi model ke data baru (Bhavsar & Panchal, 2012).

*Support Vector Machine* juga merupakan salah satu algoritma *supervised machine learning* yang banyak digunakan untuk polaritas teks (Ahmad, Aftab, & Ali, 2017).

Dalam *Support Vector Machine*, ada 3 parameter yang digunakan yaitu: *Precision*, *Recall*, *F-Measure* (Ahmad et al., 2018).

*Precision* dapat dihitung dengan menggunakan TP dan tingkat FP sebagaimana yang ditunjukkan di bawah ini:

$$Precision = \frac{TP}{(TP + FP)}$$

#### **Rumus 2.2 Rumus *Precision***

TP digunakan untuk kalimat yang diklasifikasikan dengan benar, sedangkan FP digunakan untuk kalimat yang pengklasifikasiannya salah.

*Recall* dapat dihitung dengan menggunakan rumus dibawah ini:

$$Recall = \frac{TP}{(TP + FN)}$$

#### **Rumus 2.3 Rumus *Recall***

FN digunakan untuk kalimat tidak terklasifikasi sedangkan TP adalah kalimat yang diklasifikasikan dengan benar.

F-Measure dapat dihitung dengan menggunakan rumus dibawah ini:

$$F - Measure = \frac{Precision * Recall * 2}{(Precision + Recall)}$$

#### **Rumus 2.4 Rumus *F-Measure***

### **2.8 *Decision Tree C4.5***

*Decision Tree* adalah jenis metode statistik yang dilakukan melalui representasi grafis dari proses pengambilan keputusan dalam beberapa kondisi tertentu (Kapil Mittal, Dinesh Khanduja, 2017).



*Decision Tree* juga dapat diartikan sebagai sebuah pohon di mana setiap *node* menunjukkan atribut, setiap cabang menunjukkan keputusan, dan setiap daun menunjukkan hasil (Patel & Prajapati, 2018).

Algoritma yang dipakai dalam *Decision Tree* adalah *C4.5. Decision Tree C4.5* dibangun dengan beberapa tahapan termasuk atribut sebagai root, membuat cabang untuk setiap nilai dan membagi ke cabang (Wahyuni, 2018).

## **2.9 R Programming**

R adalah bahasa pemrograman *open source* yang menggunakan data permodelan, penanganan, statistik, prediksi, waktu analisis dan visualisasi data (Prathibha, Thillainayaki, & Jenneth, 2018).

R sendiri merupakan bahasa pemrograman yang umum digunakan dalam perhitungan statistik, analisis data, dan penelitian ilmiah. R merupakan salah satu bahasa pemrograman yang paling populer untuk mengambil, menganalisis, dan menampilkan data. R menjadi populer dikarenakan sintaks ekspresinya dan antarmuka yang mudah digunakan.

## **2.10 Microsoft Power BI**

*Power BI* dikembangkan oleh *Microsoft* pada September 2013 sebagai *Power BI* untuk *Office 365*. *Power BI* pertama kali dirilis ke masyarakat umum pada 24 juli 2015.

*Microsoft Power BI* adalah alat analitik untuk menganalisis data, berbagi wawasan, dan mendapat jawaban dengan cepat dengan bantuan data interaktif visualisasi menggunakan *dashboard* yang tersedia di setiap perangkat seperti aplikasi, *dekstop*, ponsel, dll (Bhargava, Phani, Kiran, & Rao, 2018).

## 2.11 Penelitian Terdahulu

**Tabel 2.2 Tabel Penelitian Terdahulu**

No	Nama Penulis	Nama Jurnal	Judul Jurnal	Hasil Penelitian	Yang Didapat
1	Eman M.G. Younis	<i>International Journal of Computer Applications</i> (2015) Vol.112 Issue: 5 Pages: 44-48	<i>Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study</i>	Hasil penelitian ini adalah text mining dan sentimen analisis dapat dilakukan dengan menggunakan <i>open source package</i> dari aplikasi R.	Penggunaan <i>text mining</i> terhadap sentimen - sentimen di media sosial.
2	K. Prasanna Lakshmi, V. Shraddha, V. Abhinava, K. Kavya dan R. Gayathri	<i>Indian Journal of Science and Technology</i> (2017) Vol.10 Issue: 28 Pages: 1-11	<i>Sentiment Analysis and Prediction using Text Mining</i>	Hasil penelitian ini adalah mengusulkan rekomendasi model berdasarkan informasi sentimen dari ulasan yang diberikan oleh semua pengguna media sosial.	Penggunaan sentimen analisis dengan metode <i>text mining</i> terhadap media sosial.
3	M. Sarifah	<i>International Journal of Artificial Intelligent and Informatics</i> (2018) Vol.1 Issue: 2 Pages: 76	<i>Naive bayes algorithm performance for smartphone sentiment analysis in social media</i>	Hasil penelitian ini adalah pengklasifikasian sentimen analisis menjadi 3 klasifikasi yaitu positif, negatif, dan netral menggunakan algoritma Naïve Bayes.	Penggunaan algoritma Naïve Bayes di sentimen analisis.

No	Nama Penulis	Nama Jurnal	Judul Jurnal	Hasil Penelitian	Yang Didapat
4	F. Natalia, Y. Eko, F. Ferdina and et al.	<i>ICIC Express Letters, Part B: Applications</i>	<i>Interactive dashboard of flood patterns using clustering algorithms</i>	Hasil penelitian ini adalah sebuah visualisasi <i>dashboard</i> yang menunjukkan ketinggian air berdasarkan jam, bulan, dan cluster dari tahun 2013-2017.	Visualisasi terhadap suatu data yang telah diproses yang kemudian dijadikan sebuah <i>dashboar d</i>

Penelitian ini mengadopsi beberapa hal dari penelitian terdahulu. Ada 4 penelitian terdahulu yang digunakan. Penelitian pertama yang dilakukan oleh Eman M. G. Younis dengan judul “*Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study*” (Younis, 2015) menjelaskan tentang langkah-langkah yang harus dilakukan dalam melakukan *text mining*. Sehingga hal yang diambil dari penelitian ini adalah langkah-langkah yang akan dilakukan ketika melakukan *text mining* dalam penelitian ini.

Penelitian kedua yang dilakukan oleh K. Prasanna Lakshmi, V. Shraddha, V. Abhinava, K. Kayva, R. Gayathri dengan judul “*Sentiment Analysis and Prediction using Text Mining*” (Lakshmi, Shraddha, Abhinava, Kavya, & Gayathri, 2017) menjelaskan tentang melakukan analisis sentimen dalam *text mining*. Sehingga hal yang diambil dari penelitian ini adalah cara dalam melakukan analisis sentimen yang akan dilakukan ketika melakukan klasifikasi sentimen dalam penelitian ini.

Penelitian ketiga yang dilakukan oleh Monalisa Fatmawati Sarifah dengan judul “*Naive bayes algorithm performance for smartphone sentiment analysis in social media*” (Sarifah, 2018) menjelaskan tentang implementasi algoritma *naïve bayes* terhadap klasifikasi sentimen yang telah dibuat. Sehingga hal yang diambil dari penelitian ini adalah cara dalam melakukan implementasi algoritma *naïve bayes* terhadap klasifikasi sentimen.

Penelitian keempat yang dilakukan oleh Friska Natalia, Yustinus Eko, Ferry Vincettius Ferdinand, I. Made Murwantara, Chang Seong Ko dengan judul “*Interactive dashboard of flood patterns using clustering algorithms*” (Natalia, Eko, Ferdinand, Murwantara, & Ko, 2019) menjelaskan tentang *clustering* sebuah data yang kemudian dijadikan visualisasi menggunakan aplikasi Power BI. Sehingga hal yang diambil dalam penelitian ini adalah menggunakan data yang telah diperoleh kemudian dijadikan sebuah visualisasi *dashboard* dalam aplikasi Power BI.