

BAB III

METODE PENELITIAN

3.1. Bahan Penelitian

Bahan penelitian ini diambil dari Buletin APTIKOM berupa buletin sesuai tema yang telah diunggah ke aplikasi Buletin APTIKOM yang terbit setiap satu bulan sekali.

3.2. Peralatan Penelitian

Peralatan Penelitian ini menggunakan *Hardware* dan *Software*.

1. *Hardware* yang digunakan dalam penelitian ini adalah laptop dengan spesifikasi :
 - Windows 10
 - *Processor* Intel(R) Core(TM) i5-9300H CPU @ 2.40 GHz,
 - RAM 8 GB
 - SSD 512 GB.
2. Aplikasi *Software* yang digunakan dalam penelitian ini adalah :
 - Microsoft Word 2016 berfungsi untuk menyusun proposal.
 - R Studio yang merupakan aplikasi yang digunakan untuk menganalisis penelitian.

3.3. Lokasi Penelitian dan Waktu Penelitian

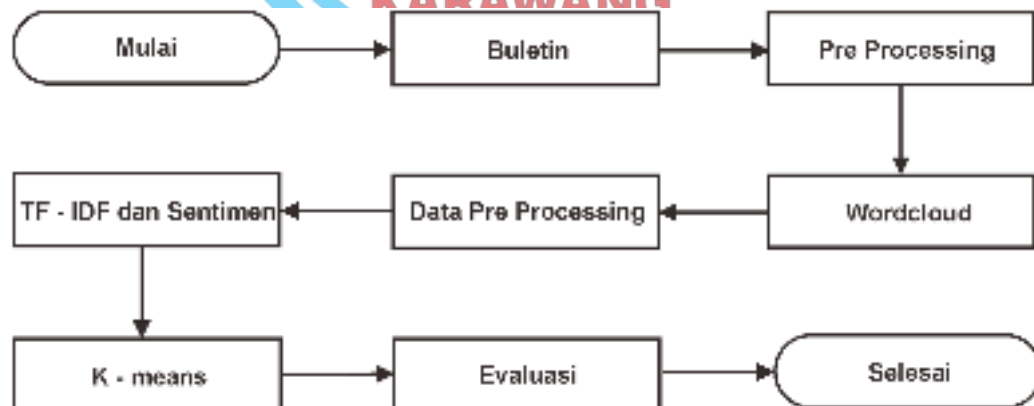
Lokasi penelitian berlokasi di Laboratorium Riset Universitas Buana Perjuangan (UBP) Karawang. Kemudian waktu penelitian berlangsung sejak bulan Januari dan rincian kegiatan penelitian ditunjukkan pada Tabel 3.1.

Tabel 3. 1 Kegiatan Penelitian

	Januari				Februari				Maret				April				Mei				Juni			
Kegiatan	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Analisis Data																								
Pre Processing																								
Implementation																								
Clustering																								
Evaluasi																								
Pengujian																								

3.4. Prosedur Penelitian

Prosedur penelitian mempunyai tahapan proses seperti pada Gambar 3.1.



Gambar 3. 1 Alur Prosedur Penelitian

Beberapa tahapan proses dalam prosedur penelitian yaitu :

1. Buletin, data berupa teks.
 2. *Pre Processing*, pemrosesan awal teks.
 3. *Word cloud*, hasil proses *pre processing* menjadi *word cloud*.
 4. *Data pre processing*, hasil proses *pre processing* menjadi dokumen.
 5. *Term Frequency – Inverse Document Frequency* (TF - IDF) dan Sentimen Analisis, pemrosesan menghitung bobot atau nilai sebuah kata terhadap suatu dokumen serta menentukan sentimen analisis.
 6. K-means, pemrosesan algoritma.
 7. Evaluasi.
- Buletin
Buletin merupakan data yang diambil dari sebuah aplikasi yang bernama Buletin APTIKOM dengan berbentuk dokumen.
 - *Pre Processing*
Pre processing salah satu tahapan menghilangkan permasalahan-permasalahan yang dapat mengganggu hasil daripada proses data. Dalam kasus klasifikasi dokumen yang menggunakan data bertipe teks. Berikut merupakan tahapan pada *pre processing* diantaranya:
 1. *To lower*, perubahan semua karakter huruf menjadi huruf kecil.
 2. *Remove numbers* merupakan penghilangan karakter angka yang ada pada kata.
 3. *Remove punctuation* merupakan penghilangan pembatas seperti tanda simbol.
 4. *Strip white space* merupakan penghilangan spasi yang berlebihan pada dokumen.
 5. *Remove words* merupakan penghilangan kata-kata pada dokumen. *Stop words* merupakan kosakata yang bukan merupakan ciri (kata unik) pada dokumen.
 6. *Stem document* merupakan proses pemetaan dan penguraian berbagai bentuk dari suatu kata menjadi bentuk kata dasar atau bisa disebut stem.

- *Word cloud*

Word cloud merupakan tahapan yang berupa hasil dari *pre processing*.

- *Data Pre Processing*

Data Pre Processing merupakan hasil pemrosesan *pre processing* yang diubah dengan berbentuk dokumen.

- *Term Frequency – Inverse Document Frequency* (TF - IDF) dan Sentimen Analisis

TF - IDF ini adalah langkah sesudah *pre processing* yang sudah diproses menjadi data bersih dari sebuah dokumen. Cara kerja dari TF – IDF ini merupakan teknik untuk menghitung bobot atau nilai sebuah kata terhadap suatu dokumen dan menentukan sentimen berupa kalimat positif dan negatif.

- K-means

K-means digunakan pada penelitian ini adalah untuk mengetahui nilai *cluster* atau kelompok dalam kalimat positif maupun negatif.

- *Evaluation*

Evaluasi pada penelitian ini adalah dengan menggunakan metode SSE atau bisa disebut dengan *Sum Square Error*. *Sum Square Error* merupakan cara dalam melakukan validasi *cluster* melalui jumlah kuadrat setiap anggota *cluster* menuju pusatnya (Duvvada, 2017).

3.5. Implementasi

Penelitian ini dilakukan untuk mengimplementasikan dan menganalisis sentimen terhadap penulisan Buletin APTIKOM. Bertujuan mengelompokkan kalimat positif negatif dan menentukan kalimat positif maupun negatif dengan menggunakan algoritma K-Means.

Penelitian ini dilakukan dengan pengimplementasian menggunakan R Studio atau bahasa pemrograman R dengan penggunaan *text processing*, metode TF-IDF (*Terms Frequency-Inverse Document Frequency*) untuk menentukan bobot nilai pada dokumen dan SSE (*Sum Square Error*) untuk evaluasi. Data yang diambil berupa dokumen buletin APTIKOM. Serta penggunaan algoritma K-Means yang akan menentukan *cluster* atau kelompok kalimat atau kata positif maupun negatif. Metode TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen (Nurjannah & Fitri Astuti, 2013). Persamaan (4) merupakan cara untuk TF-IDF:

$$tf = 0,5 + 0,5x \frac{tf}{\max (tf)}$$

$$idf_t = \log \left(\frac{D}{df_t} \right)$$

$$W_{d,t} = tf_{d,t} \times IDF_{d,t}$$
(4)

Keterangan :

tf = banyaknya kata yang dicari pada sebuah dokumen

max(tf) = jumlah kemunculan terbanyak term pada dokumen yang sama

Nilai D = total dokumen

df_t = jumlah dokumen yang mengandung term t.

IDF = *Inversed Document Frequency* ($\log_2(D/df)$)

d = dokumen ke d

t = kata ke-t dari kata kunci

W = bobot dokumen ke-d terhadap kata ke-t

3.6. Analisis Data

Menganalisis data berupa dokumen untuk menentukan sentimen pada penulisan dalam *clustering* kata yang termasuk kata positif maupun negatif.