BAB III

METODOLOGI PENELITIAN

1.1. Bahan Penelitian

Data yang digunakan bahan penelitian sentimen yaitu *magazine* Buletin yang diterbitkan oleh APTIKOM berisi pengetahuan teknologi dan informasi setiap bulan.

3.1. Peralatan Penelitian

Alat yang digunakan penelitian sentimen yaitu *software Rstudio*, *Microsoft Word* 2010, *Microsoft Excel* dan *Hardware* yaitu laptop dengan spesifikasi AMD *Ryzen* 5 2500U *with Radeon Vega Mobile*, 1 *Terabyte*, 8 GB RAM, *mouse*, *flashdisk*.

3.2. Waktu dan Lokasi Penelitian

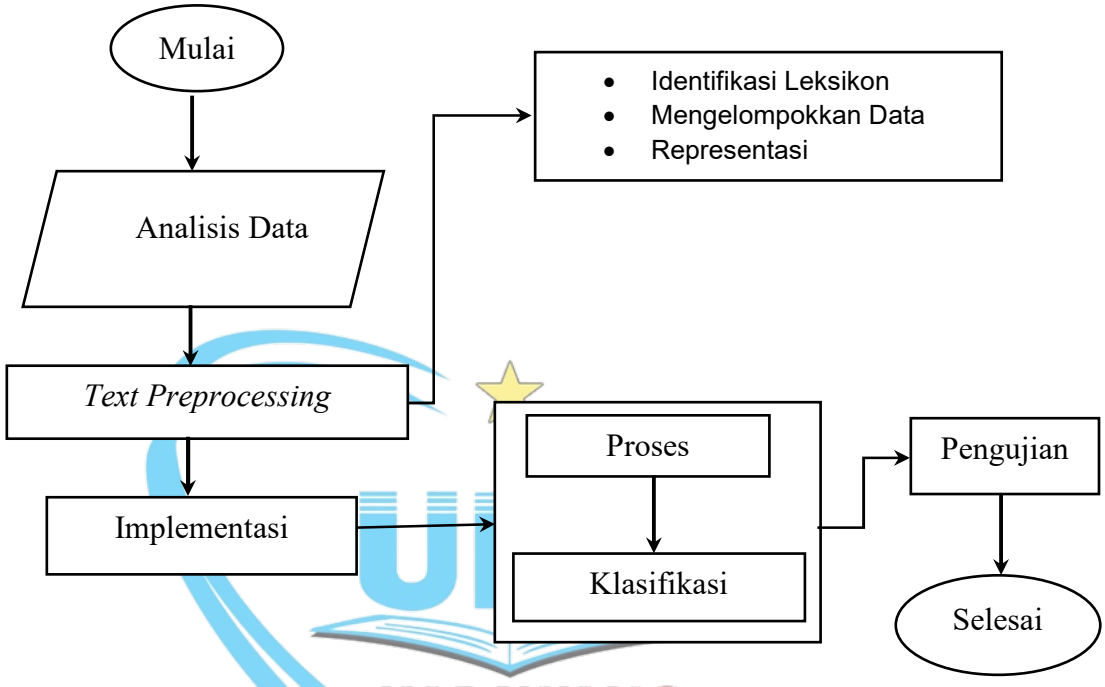
Lokasi pada kegiatan penelitian dilaksanakan di Laboratorium Riset Universitas Buana Perjuangan Karawang dilaksanakan sejak bulan November sampai bulan Juli. Detail pelaksanaan penelitian ditunjukkan pada Tabel 3.1.

Tabel 3. 1 Tabel Penelitian Terkait

November 2020	Desember	Januari	Februari	Maret	April	Mei	Juni	Juli
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**	**1 2 3 4**	**1 2 3 4**
Studi Literatur								
Penulisan Proposal								
Analisis Data *Preprocessing*								
Implementasi								
Klasifikasi								
Evaluasi								
Pengujian								
Penulisan Laporan Tugas Akhir								
Yudisium								

3.3. Prosedur Penelitian

Tahapan prosedur penelitian dimulai dengan melakukan analisis data pada *magazine* Buletin untuk mengetahui fakta dan data sebenarnya. Prosedur penelitian dalam melakukan analisis sentimen ditunjukkan pada Gambar 3.1.



Gambar 3. 1 Gambar Prosedur Penelitian

1. Analisis data dilakukan untuk mengetahui permasalahan yang terdapat pada *magazine* Buletin dengan mengambil beberapa sampel kalimat yang tertera pada *magazine* Buletin.
2. *Text Preprocessing* merupakan tahapan untuk mengetahui aktualitas, kelengkapan data dengan melakukan identifikasi leksikon, mengelompokkan data, dan melakukan representasi data.
3. Implementasi algoritme dimulai dengan melakukan, proses perbandingan nilai, dan pengelompokkan sentimen.
4. Pengujian dilakukan untuk mengetahui tingkat akurasi algoritme menggunakan metode *Sum of Square Error* (SSE) untuk menguji performa algoritme DBSCAN.

1.5. Analisis Data (Permasalahan)

Pada tahap analisis data, proses pencarian permasalahan data dilakukan dengan tujuan untuk mengetahui fakta data sebenarnya. Cara untuk menentukan fakta data sentimen bernilai positif dan negatif dapat digambarkan sebagai berikut:

1. Sentimen positif dapat diartikan sebagai suatu tindakan yang dapat dilakukan secara pasti dan tidak adanya penolakkan.
2. Sentimen negatif dapat diartikan sebagai suatu tindakan yang belum diketahui secara pasti dan bersifat kurang baik.

Standar yang digunakan untuk analisis sentimen yaitu berdasarkan *SentiWordNet* yang menjadi sumber leksikal dalam melakukan penambangan teks. SentiWordNet memiliki tiga kategori pada setiap *synset* di antaranya positif, negatif dan skor objektivitas (M, S, & I.K.E, 2017). Setiap *synset* memiliki kategori nilai dengan skor lebih besar dari 0 bersifat positif dan kurang dari 0 bersifat negatif (Han & Zhang, 2018). Tahapan dalam proses analisis data di antaranya mengumpulkan data, identifikasi sentimen, menentukan metode dan menafsirkan hasil analisis. Tujuan dari proses analisis sentimen yaitu untuk mendapatkan data yang dapat diterjemahkan ke dalam bahasa yang lebih mudah. Selain itu, analisis data dilakukan untuk mendapatkan kesimpulan dari data yang telah dianalisis berdasarkan hasil pengujian hipotesis.

1.6. *Text Preprocessing*

Pada tahap *text processing*, informasi yang telah diperoleh akan dilakukan proses penyeleksian data dengan tujuan menjadikan data tidak terstruktur menjadi data terstruktur. *Text preprocessing* dilakukan untuk menentukan kata yang akan dijadikan sebagai parameter sentimen. Parameter yang telah ditentukan merupakan kata – kata yang mewakili isi dari dokumen maupun lainnya. Terdapat beberapa tahapan untuk melakukan *text processing* di antaranya:

1. Identifikasi Leksikon

   Pada identifikasi leksikon, proses penyaringan data dilakukan untuk menetapkan parameter dari sebuah leksikon. Setelah itu, proses penyaringan leksikon akan kategorikan sesuai dengan parameter sentimen. Sentimen yang diperoleh akan dijadikan sebagai data rujukan penambangan teks. Proses identifikasi akan menghasilkan kandidat yang akan ditetapkan sebagai parameter sentimen.

2. Mengelompokkan Data

   Pengelompokkan data dilakukan untuk mengetahui dan membedakan setiap leksikon yang akan didistribusikan sesuai parameter sentimen. Pada tahap pengelompokkan data, data yang telah diperoleh akan dikelompokkan ke dalam sentimen positif dan negatif.

3. Representasi

   Representasi merupakan tahapan interpretasi leksikon untuk menentukan dan memverifikasi kesesuaian dengan parameter sentimen. Sentimen yang sesuai akan diproses untuk dilakukan proses pengujian dan akurasi.

1.7. Implementasi Algoritme

Penerapan algoritme diawali dengan menerapkan algoritme DBSCAN untuk melakukan *clustering* pada sebaran data. Proses *clustering* dilakukan dengan menentukan radius dan minimum poin pada sebaran data. Poin yang termasuk ke dalam jangkauan radius dan minimum poin. Cara untuk mendapatkan pola sering muncul dan klasifikasi dari sebaran data maka menggunakan formula TF – IDF. Berikut beberapa tahapan yang dilakukan:

1. Proses

   Proses implementasi leksikon akan menerapkan formula TF – IDF dalam melakukan penambangan teks. TF – IDF akan menghitung nilai perbandingan setiap leksikon untuk mendapatkan kata yang sering digunakan dalam dokumen. Tujuan proses TF – IDF yaitu menghasilkan nilai yang telah

dikalkulasi dan menemukan pola sering muncul. Persamaan (2), (3), dan (4) merupakan cara untuk membentuk TF – IDF (Isnarwaty & Irhamah, 2019).

$$TF_j = \left(\frac{i}{DF_j}\right) \quad 1.$$

$$IDF_j = log\ log\ \left(\frac{N}{DF_j}\right) \quad (3)$$

$$W_{i,j} = TF_{i,j} \times IDF_j, \quad (4)$$

Keterangan:

i = *term*

j = dokumen

$W_{i,j}$ = bobot dari kata ke j pada kata i

IDFj = *inverse document frequency* pada kata ke j

TF$i,j$ = jumlah kemunculan kata ke j pada kata ke i

DFj = banyaknya kalimat yang mengandung kata j

N = jumlah keseluruhan kata

2. Klasifikasi

   Klasifikasi merupakan salah satu cara untuk mengetahui dan melakukan pengelompokkan *itemset* berdasarkan bobot nilai setiap *itemset*. Klasifikasi dapat dilakukan dengan pengumpulan data untuk mencari kata tertentu dari dokumen.

3. Evaluasi

   Evaluasi merupakan salah satu cara untuk mengetahui kesesuaian dan kebenaran dari data analisis. Proses evaluasi akan memeriksa kesesuaian kata dengan kelas yang didistribusikan pada sentimen. Tujuan dari evaluasi yaitu untuk mengetahui tingkat kesesuaian kata dengan kelas sentimen.

1.8. Pengujian

Pada tahap ini, proses pengujian dilakukan untuk memastikan aktualitas dari proses pencarian *frequent item*. Klasifikasi sentimen menggunakan formula TF – IDF kemudian hasilnya diolah menggunakan DBSCAN. Hasil yang didapatkan kemudian

dievaluasi menggunakan rumus *Sum of Square Error* (SSE). SSE dapat menghitung perbandingan dari masing – masing nilai *cluster*. Persamaan (5) merupakan cara untuk menghitung *Sum of Square Error* (SSE) (Mar'i & Supianto, 2018).

$$SSE = \sum_{K=1}^{K} \sum_{x_i \in S_k} \|X_i - C_j\|22 \quad (4)$$

Keterangan:

Xi = Data ke - *i*

Cj = Titik pusat *cluster* atau (*centroid*)

K = Jumlah *cluster*

Setelah melakukan evaluasi dengan SSE maka nilai yang didapatkan akan dilakukan pemeriksaan *cluster* dengan *Silhouette Coefficient*. Tujuan dari pemeriksaan yaitu untuk menilai kualitas *cluster* yang dihasilkan. Persamaan (6) adalah *Silhouette Coefficient* untuk menghitung ketepatan pada pengelompokkan *cluster* (Saputra & Chusyairi, 2020).

$$S_i = \frac{(b_i - a_i)}{max\,(b_i - a_i)} \quad (5)$$

Keterangan:

a_i = nilai rata – rata jarak objek i dengan seluruh objek pada *cluster*

b_i = nilai terkecil rata – rata objek dalam *cluster*

S_i = hasil rata – rata *cluster*


UBP
KARAWANG