

BAB III METODE PENELITIAN

3.1. Bahan Penelitian

Bahan yang digunakan dalam penelitian ini adalah dataset hasil penambangan data dari *website Twitter*. Dataset yang digunakan berupa kalimat-kalimat tentang konflik sosial yang ditulis oleh pengguna.

3.2. Peralatan Penelitian

Perangkat keras yang digunakan adalah Processor AMD Ryzen 5 2700x (3.7 GHz), Motherboard Gigabyte AB-350-Gaming, RAM Team Elite 16 GB, Hardisk Visipro 120 GB, VGA NVidia GTX 1050 4GB, Monitor Dell 21” Resolusi 1920x1024, Keyboard dan Mouse Logitech. Bahasa pemrograman untuk melakukan analisis dan pembangunan model menggunakan *R Studio*.

3.3. Pelaksanaan Penelitian

Lokasi penelitian dilaksanakan di Laboratorium Riset UBP Karawang terhitung sejak bulan Januari 2022 dengan rincian pelaksanaan penelitian ditampilkan pada Tabel 3.1.

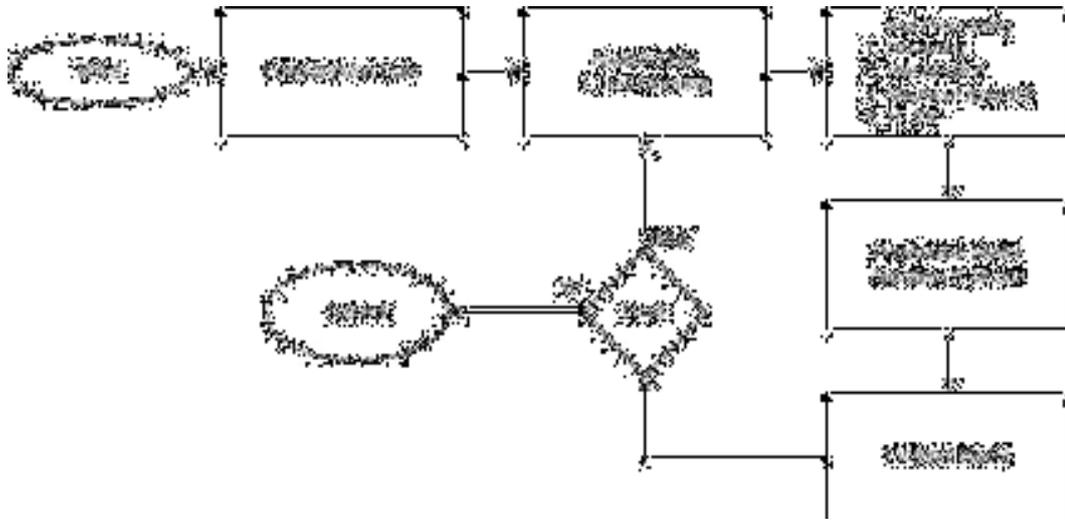
Tabel 3. 1 Rincian Penelitian

No	Item	2022						
		1	2	3	4	5	6	7
1	Pengumpulan data							
2	Analisis data							
3	Proses <i>text mining</i>							
4	Pembuatan Model algoritma <i>K-Means</i>							
5	Evaluasi							
6	Hasil							

3.4. Prosedur Penelitian

Kegiatan pertama yang dilakukan dalam penelitian ini adalah menganalisis dan mengumpulkan data, kedua melakukan perancangan kode, ketiga

mengelompokkan data yang didapatkan lalu dilakukan evaluasi model. Gambar 3.1 merupakan *Flowchart* prosedur penelitian.



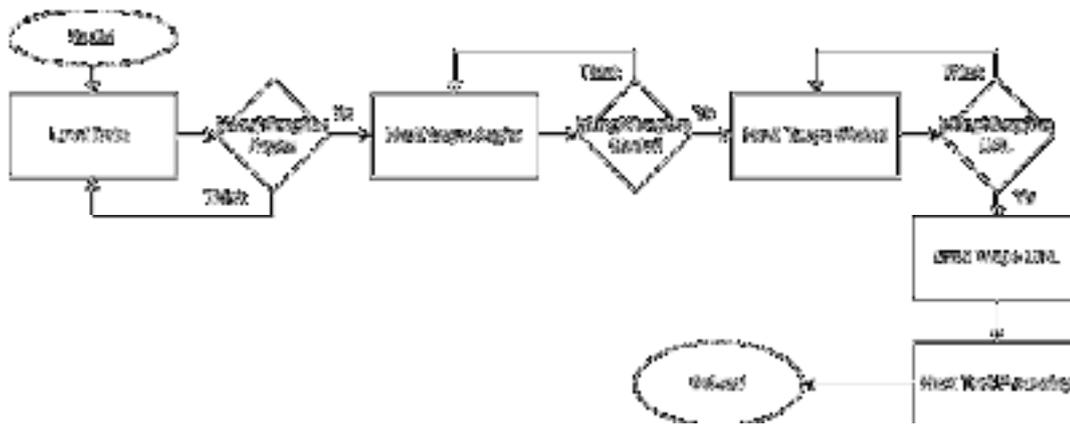
Gambar 3. 1 Prosedur Penelitian

3.5. Pengumpulan Data

Penelitian ini dimulai dengan pengumpulan data dengan cara menambang data melalui media sosial *twitter*. Data hasil penambangan berjumlah 3952 *twit* dengan waktu penambangan dari Bulan Mei sampai Juni 2022.

3.6. Analisis Data

Analisis data dilakukan untuk mengetahui topik dari data yang didapatkan dengan kata kunci konflik sosial. Data-data yang telah dianalisis akan mendapatkan hasil yaitu informasi topik-topik terkait dengan konflik sosial yang sedang terjadi di masyarakat Indonesia. Kegiatan yang dilakukan dalam proses ini adalah memilih data dan membersihkannya menggunakan metode *text cleaning*. Proses ini ditunjukkan pada Gambar 3.2.

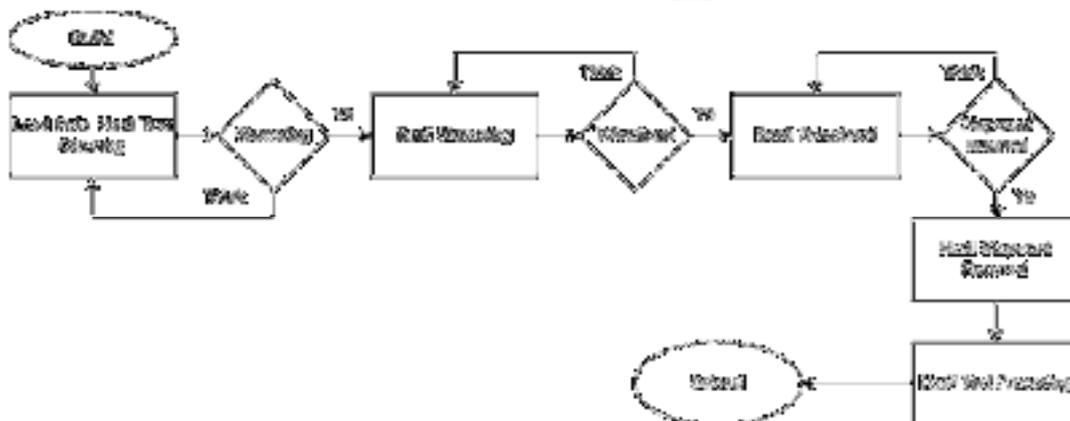


Gambar 3. 2 *Flowchart Text Cleaning*

Proses *text cleaning* dilakukan menggunakan bahasa *R* yang ditampilkan pada tabel 3.2.

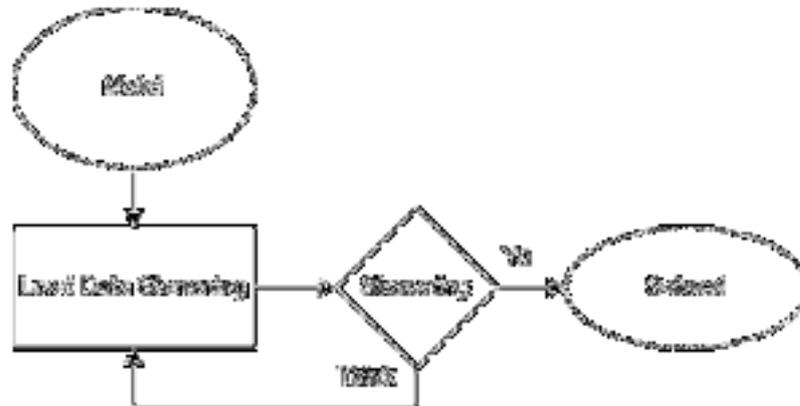
3.7. Text Mining

Hasil informasi yang didapatkan dari analisis data diolah menggunakan teknik *Text Mining* untuk menglempokan barisan teks. Tahapan proses dimulai dengan *Text Processing* yang didalamnya terdapat empat kegiatan. Kegiatan satu yaitu *Stemming* yang berfungsi merubah kata berimbuhan menjadi kata dasar. Kedua adalah *Tokenization* untuk memecah kalimat menjadi kata-kata. Ketiga menggunakan teknik *Stopword Removal* yang bertujuan menghilangkan kata-kata yang tidak berarti. Keempat melakukan pembobotan kata dengan menggunakan algoritma *TF-IDF* agar mendapatkan angka frekuensi keseringan kata muncul. Proses ini ditampilkan pada Gambar 3.3.



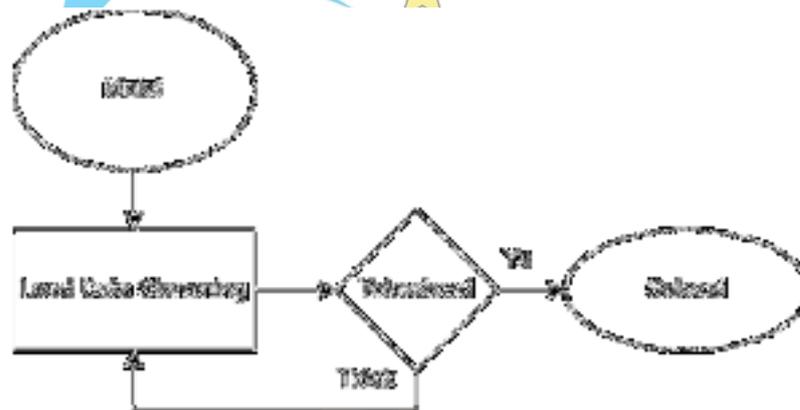
Gambar 3. 3 *Flowchart Text Processing*

Berikut adalah Gambar 3.4 yang merupakan alur kerja dari proses *stemming* untuk menghilangkan angka, symbol dan alamat *url*.



Gambar 3. 4 *Flowchart* Stemming

Selanjutnya adalah Gambar 3.5 yang merupakan alur kerja dari proses *tokenizing* untuk menghilangkan merubah kalimat menjadi kata-kata.



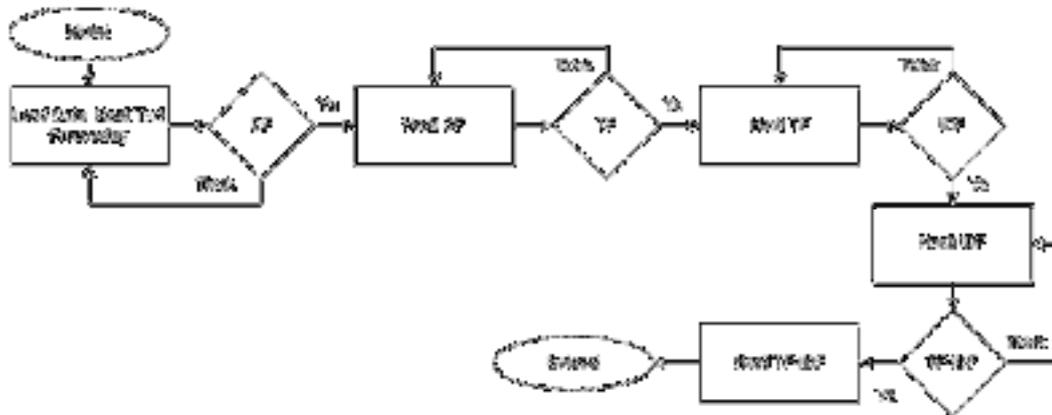
Gambar 3. 5 *Flowchart* Tekonisasi

Lalu pada Gambar 3.6 merupakan alur kerja dari proses *stopword removal* untuk menghilangkan kata-kata yang tidak berhubungan dengan topik.



Gambar 3. 6 *Flowchart* Stopword Removal

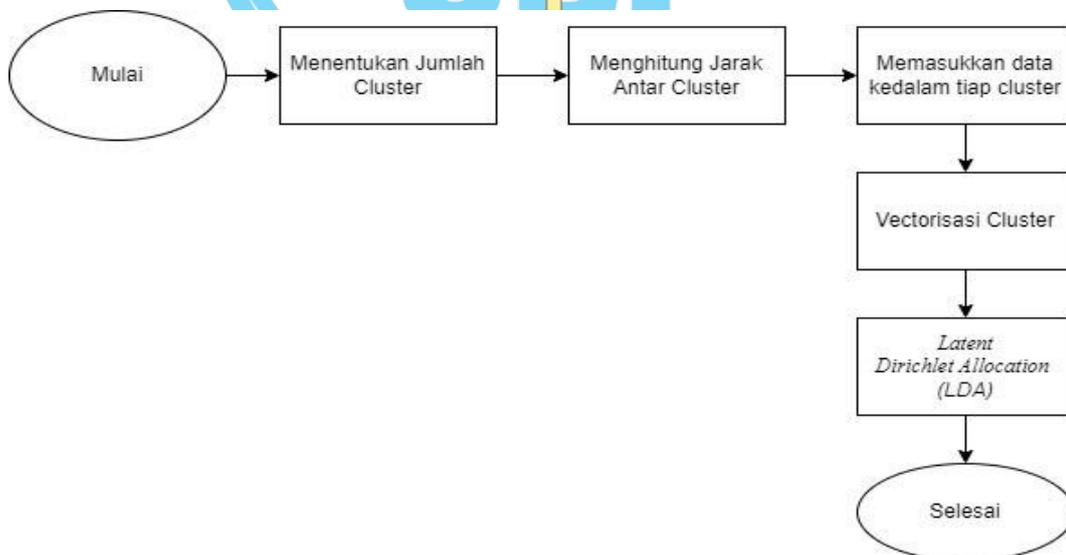
Proses terakhir adalah *TF-IDF* untuk mengetahui keseringan kata muncul dengan menghitung bobot dari tiap kata dan menampilkan hasil menggunakan fitur wordcloud. Gambar 3.7 merupakan alur proses dan metode *TF-IDF*.



Gambar 3. 7 Flowchart TF-IDF

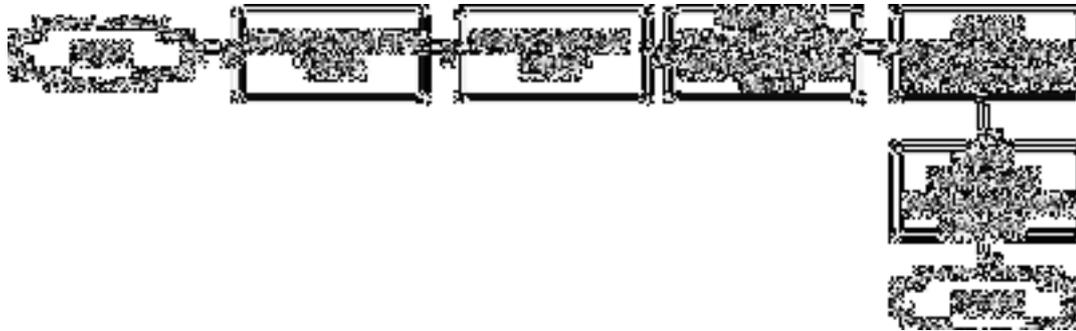
3.8. Pembuatan Model

Model akan dibuat dengan menggunakan algoritma *K-Means* dengan melalui langkah-langkah sebelumnya. Proses pembuatan model menggunakan metode *K-Means* ditampilkan pada Gambar 3.8.



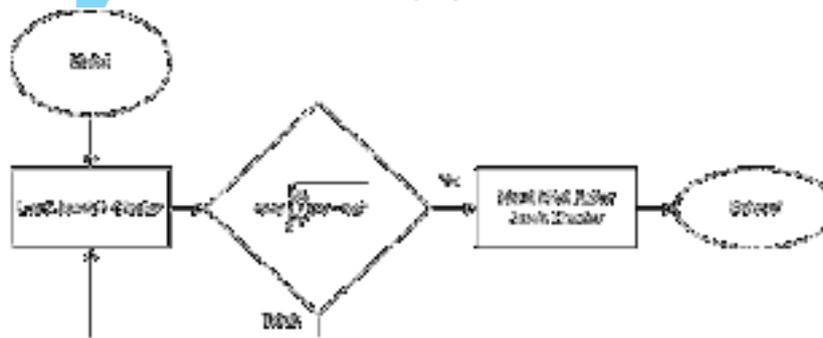
Gambar 3. 8 Flowchart K-Means

Proses pertama yaitu menentukan jumlah *cluster* menggunakan metode *Elbow Plot* yang disajikan pada Gambar 3.5 berikut ini :



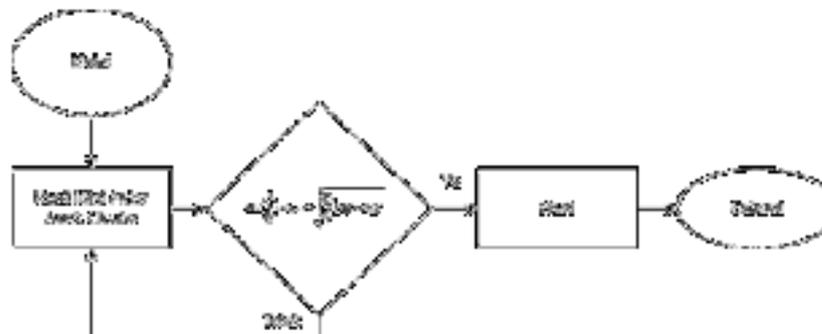
Gambar 3. 9 *Flowchart Elbow Method*

Selanjutnya proses menghitung jarak antar *cluster* menggunakan metode *sillhoute coeffisien* yang ditampilkan pada Gambar 3.10.



Gambar 3. 10 *Flowchart Jarak Antar Cluster*

Berikutnya adalah memasukan seluruh anggota data kedalam tiap *cluster* yang ditampilkan pada Gambar 3.11.



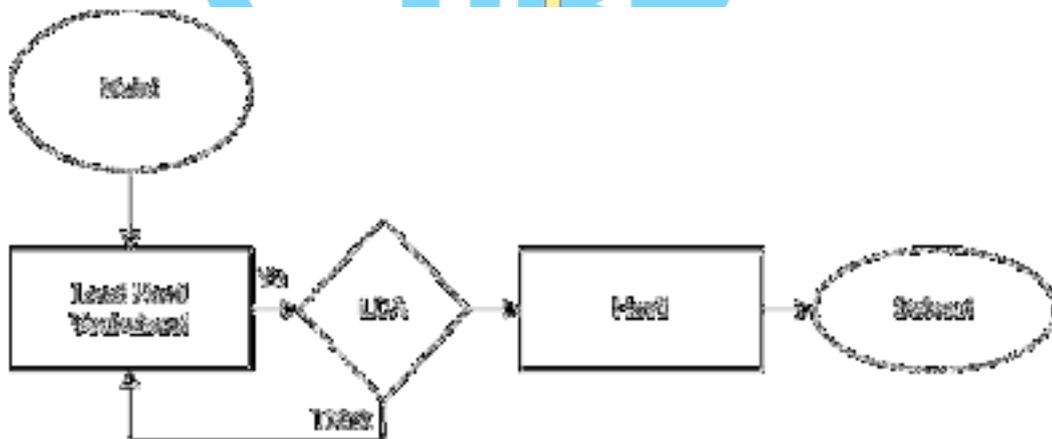
Gambar 3. 11 *Flowchart Memasukan Anggota*

Setelah memasukan anggota data ketiap cluster maka dilakukan *vectorisasi cluster* untuk memilih kata-kata yang sesuai. Berikut ini adalah Gambar 3.12 yang berisikan proses *vectorisasi cluster*.



Gambar 3.12 Flowchart Vectorisasi Cluster

Proses terakhir yaitu melihat keterangan cluster yang dihasilkan menggunakan metode *Latent Dirichlet Allocation* yang ditampilkan pada Gambar 3.13.



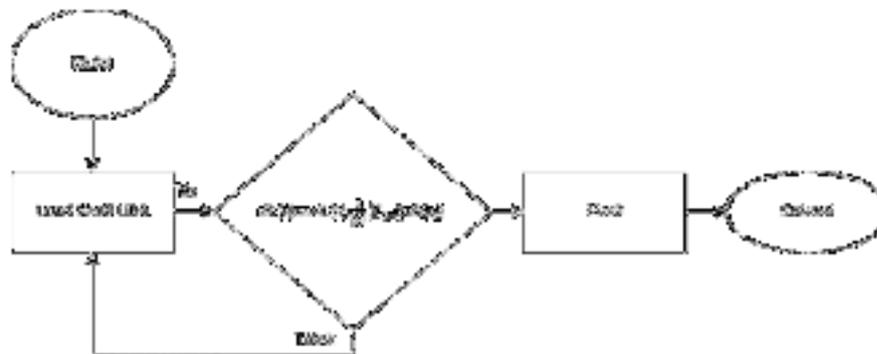
Gambar 3.13 Flowchart LDA

3.9. Evaluasi

Penelitian ini pada evaluasi menggunakan metode *Sum of Square Error* dan *Sillhoute Coeffisien*. Metode ini akan menunjukkan hasil berupa tingkat akurasi dari pemodelan yang telah dilakukan. Gambar 3.14 adalah alur kerja evaluasi *SSE* dan Gambar 3.15 adalah evaluasi *Sillhoute Coeffisien*.



Gambar 3. 14 *Flowchart Sum of Square Error*



Gambar 3. 15 *Flowchart Silhouette Coefficient*

