

BAB III

METODOLOGI PENELITIAN

3.1. Subyek dan Obyek Penelitian

Subjek yang ada dalam penelitian ini adalah kumpulan komentar atau *tweet* yang di *crawling* dari media sosial *Twitter*. Data dikumpulkan menggunakan API yang disediakan oleh *Twitter* dengan *Twitter API*, Data tersebut dikumpulkan melalui proses yang disebut sebagai *crawling* dengan *keyword* salah satu nama anggota kepolisian, karena *keyword* yang disebutkan berhubungan dengan kasus tersebut. *Keyword* yang disebutkan di atas pernah menjadi tren dalam topik Indonesia di *Twitter*. Sedangkan objek penelitian ini merupakan sentimen positif, netral, dan negatif yang terkandung dalam *tweets* publik terhadap kasus viral yang membawa nama baik kepolisian Indonesia. Gambar 3.1 merupakan *keyword* kasus salah satu oknum petinggi polisi yang sedang *trending* di sosial media *Twitter*.



Gambar 3.1 Keyword Nama Anggota Kepolisian Yang Sedang Trending di Indonesia Dengan Lebih Dari 81.000 Tweet

3.2. Alat Dan Bahan Penelitian

3.2.1. Alat Penelitian

Penelitian ini dilakukan dengan menggunakan alat berupa *hardware* atau perangkat keras dan *software* atau perangkat lunak sebagai alat yang mendukung proses pengerjaan penelitian.

1. Perangkat Keras (*Hardware*)
 - a. Asus VivoBook A442U dengan spesifikasi Processor Intel(R) Core(TM) i5-8250U 8th Gen, Ram 4GB dan Hard disk 1TB.
2. Perangkat Lunak (*Software*)
 - a. OS Windows 10
 - b. Microsof Office 2019
 - c. *Software RapidMiner Studio* dengan versi 9.9.0.0 untuk melakukan *crawling* data *Twitter*
 - d. *Software Orange Data Mining*
 - e. Microsoft Excel sebagai alat untuk penyimpanan data hasil *crawling*
 - f. Mendeley Dekstop versi 1.19.8 dan Mendeley Reference Manager versi 2.65.0 untuk mempermudah mengelola dan mencatat referensi yang digunakan

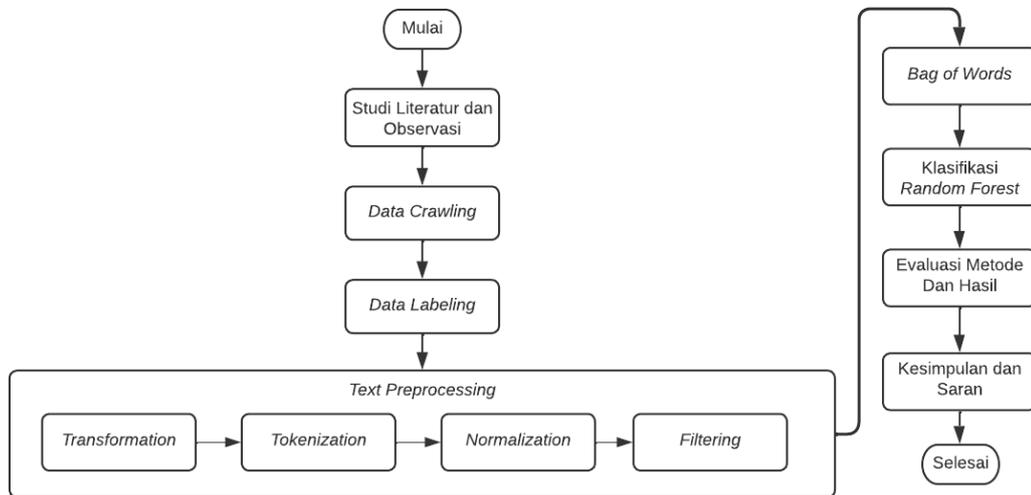
3.2.2. Bahan Penelitian

Bahan yang digunakan dalam penelitian merupakan data *Tweet* atau komentar masyarakat yang berhubungan dengan kasus viral kepolisian dengan *keyword* nama salah satu dari anggota kepolisian yang viral di *Twitter*. Data tersebut dikumpulkan melalui *Twitter API* dengan *software RapidMiner* karena penggunaannya yang mudah dan memiliki kemampuan untuk melakukan *crawling* data dengan jumlah yang banyak.

3.3. Diagram Alir Penelitian

Tahapan penelitian ditampilkan dalam diagram alir penelitian yang dilaksanakan dimulai dari tahap studi literatur hingga pada tahap akhir yaitu

memberikan kesimpulan dan saran. Berikut merupakan tahap alur penelitian ini pada gambar 3.2.



Gambar 3.2 Diagram Alir Penelitian

3.3.1. Studi Literatur dan Observasi

Studi literatur merupakan pengumpulan data pustaka seperti penelitian-penelitian terdahulu yang memiliki keterkaitan dengan penelitian ini lalu dijadikan sebagai acuan untuk menganalisis sentimen opini publik terhadap kasus viral institusi kepolisian pada *Twitter* dengan metode *Random Forest*. Tahapan ini dilakukan untuk menunjang pencapaian tujuan akhir penelitian serta sebagai sumber referensi sebelum menjalankan penelitian. Observasi dilakukan untuk mengamati opini publik di sosial media *Twitter* yang berinteraksi terhadap kasus viral tersebut.

3.3.2. Data Crawling

Data yang digunakan dan diolah dalam penelitian ini berupa data *tweets* yang mengandung *keyword* nama salah satu petinggi kepolisian yang menjadi topik pembicaraan yang tren di sosial media *Twitter*. *Crawling* data dari *Twitter*, memerlukan koneksi dari *software RapidMiner* terhadap *Twitter API (Application Programming Interface)*. Jumlah data *tweets* yang diambil

yaitu berjumlah 10.000 data, data tersebut diambil saat kasus sedang viral dan ramai menjadi perbincangan publik. Data dari hasil *crawling* ini di ekstrak ke dalam format *file .xlsx* dan digunakan dalam tahapan *preprocessing data*.

3.3.3. Labeling Data

Tahap *Labeling Data* atau pelabelan data membantu mesin untuk mempelajari pola dari data yang sudah dilabeli dan membangun model prediksi dengan menggunakan data tersebut sebagai acuan. Pada data yang sudah diambil pada tahap sebelumnya *tweet* masyarakat yang berkaitan dengan kasus diberi label positif, netral atau negatif, tergantung dari sentimen yang terkandung di dalamnya. Pelabelan dilakukan menggunakan *tools* yang ada di software *Orange Data Mining* yaitu *sentiment analysis*.

3.3.4. Data Preprocessing

Data Preprocessing merupakan tahap yang penting digunakan untuk sebuah pemrosesan data karena hal tersebut dapat mempengaruhi data yang akan digunakan, hasil yang didapatkan juga sangat berpengaruh pada *preprocessing* supaya data bisa dibaca dengan baik oleh sistem. Kegiatan yang ada pada tahap *Preprocessing* antara lain *Transformation*, *Tokenization*, *Normalization*, dan *Filtering*

1. Transformation

Transformation merupakan tahap awal pada *Data Preprocessing*, pada tahapan ini semua *tweets* yang memiliki huruf kapital (*uppercase*) diubah menjadi huruf kecil (*lowercase*), dan menghilangkan URLS yang terkandung di dalam setiap *tweet* agar memudahkan sistem untuk melakukan pemrosesan data.

2. Tokenization

Tokenization merupakan tahapan yang dilakukan dengan memotong sebuah kalimat menjadi kata perkata yang disebut sebagai token, seperti contoh “Sepertinya di Purwokerto besok akan turun hujan” lalu dipotong menjadi “Sepertinya”, “di”, “Purwokerto”, “besok”, “akan”, “Turun”, “Hujan”.

3. *Normalization*

Tahap *normalization* pada *Orange Data Mining* dapat disebut juga dengan tahap *stemming* yang merupakan salah satu tahap dalam pemrosesan teks (*text preprocessing*) yang bertujuan untuk menghasilkan bentuk dasar kata (*root word*) dari sebuah kata. *Stemming* dilakukan dengan cara menghilangkan awalan (*prefix*) dan akhiran (*suffix*) pada kata sehingga hanya menyisakan kata dasarnya.

Stemming penting dilakukan dalam pemrosesan teks karena dapat membantu mengurangi jumlah variasi kata yang digunakan, sehingga memudahkan dalam analisis data teks. Dengan melakukan *stemming*, kata-kata yang memiliki bentuk yang sama dapat dianggap sebagai satu kata, sehingga dapat menghasilkan representasi teks yang lebih sederhana dan mudah diproses. Namun, perlu diperhatikan bahwa *stemming* juga dapat menghasilkan bentuk dasar yang tidak selalu benar atau masih ambigu, terutama dalam bahasa yang memiliki banyak variasi kata seperti bahasa Indonesia.

4. *Filtering*

Filtering merupakan tahapan untuk menghilangkan kata-kata yang mempresentasikan isi data dengan membuang setiap kata yang dianggap tidak penting yang dapat disebut dengan *stopwords*. *Stopwords* dapat berupa kata sambung, kata depan, dan kata seru seperti “di”, “yang”, “dan”, “ke”, “wah”, dan lain-lain.

3.3.5. *Bag of Words*

Pada tahapan *Bag of Words* bertujuan untuk mengkonversi kata dalam vektor dan digunakan dalam beberapa aplikasi NLP (*Natural Language Processing*). Di tahap ini juga mengekstraksi fitur yang disederhanakan untuk data teks yang mudah diimplementasikan. Tahapan ini mempertahankan kosakata dan menghitung frekuensi kata, mengabaikan berbagai abstraksi bahasa alami seperti tata bahasa dan urutan kata. Pendekatan *Bag of Words* mengambil dokumen sebagai input dan memecahnya menjadi kata-kata.

3.3.6. Klasifikasi *Random Forest*

Pada tahap ini dilakukan proses klasifikasi dengan *Algoritma Random Forest*. Di tahapan ini model *Random Forest* diajari untuk mengenal dokumen yang sudah ada kedalam tiga kelas, yaitu kelas negatif, kelas netral dan kelas positif. Jumlah *Tree* yang di implementasikan yaitu sebesar 100 *Tree*, karena semakin besar jumlah *Tree* maka semakin baik hasil klasifikasi yang didapatkan.

3.3.7. Evaluasi Metode dan Hasil

Output yang dihasilkan dari metode yang digunakan akan menghasilkan model evaluasi seperti *Accuracy*, *Precision*, *F1-Score* dan *Recall*. Untuk mengetahui hasil dari uji coba metode klasifikasi algoritma *random forest* yang dipilih, pada penelitian ini metode evaluasi yang dipilih menggunakan *Confusion Matrix* metode ini merupakan salah satu metode yang digunakan dalam melakukan perhitungan akurasi pada konsep *data mining*. *Confusion Matrix* digambarkan dengan tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan. Karena penelitian ini mengklasifikasikan tiga kelas sentimen maka akan memuat sembilan kejadian ketika proses pengujian kelas seperti pada tabel berikut.

Tabel 3.1 *Confusion Matrix* dengan tiga kelas

Actual Label	Predicted Label		
	Positive (+)	Neutral (0)	Negatives (-)
Positive (+)	True Positives(TP)	False Positive (FP)	False Positive (FP)
Neutral (0)	False Neutral (Fnet)	True Neutral (TNet)	False Neutral (FNet)
Negative (-)	False Negative (FN)	False Negative (FN)	True Negative (TN)

1. *True Positive*: merupakan jumlah *record* data positif yang di klasifikasikan sebagai nilai positif.
2. *False Positive*: merupakan jumlah *record* data negatif yang diklasifikasikan sebagai nilai salah positif.

3. *False Neutral* : merupakan jumlah *record* data netral yang diklasifikasikan sebagai nilai salah netral.
4. *True Netral* : merupakan jumlah *record* data netral yang diklasifikasikan sebagai nilai benar netral.
5. *False Negatives*: merupakan jumlah *record* data positif yang diklasifikasikan sebagai nilai salah negatif.
6. *True Negatives*: merupakan jumlah *record* data negatif yang diklasifikasikan sebagai nilai benar negatif.

3.3.8. Kesimpulan dan Saran

Tahapan yang terakhir yaitu kesimpulan dan saran. Tahap ini dilakukan dengan tujuan untuk mengetahui hasil analisis sentimen opini publik terhadap kasus viral seorang anggota kepolisian dengan algoritma *Random Forest* menggunakan *software Orange Data Mining*.