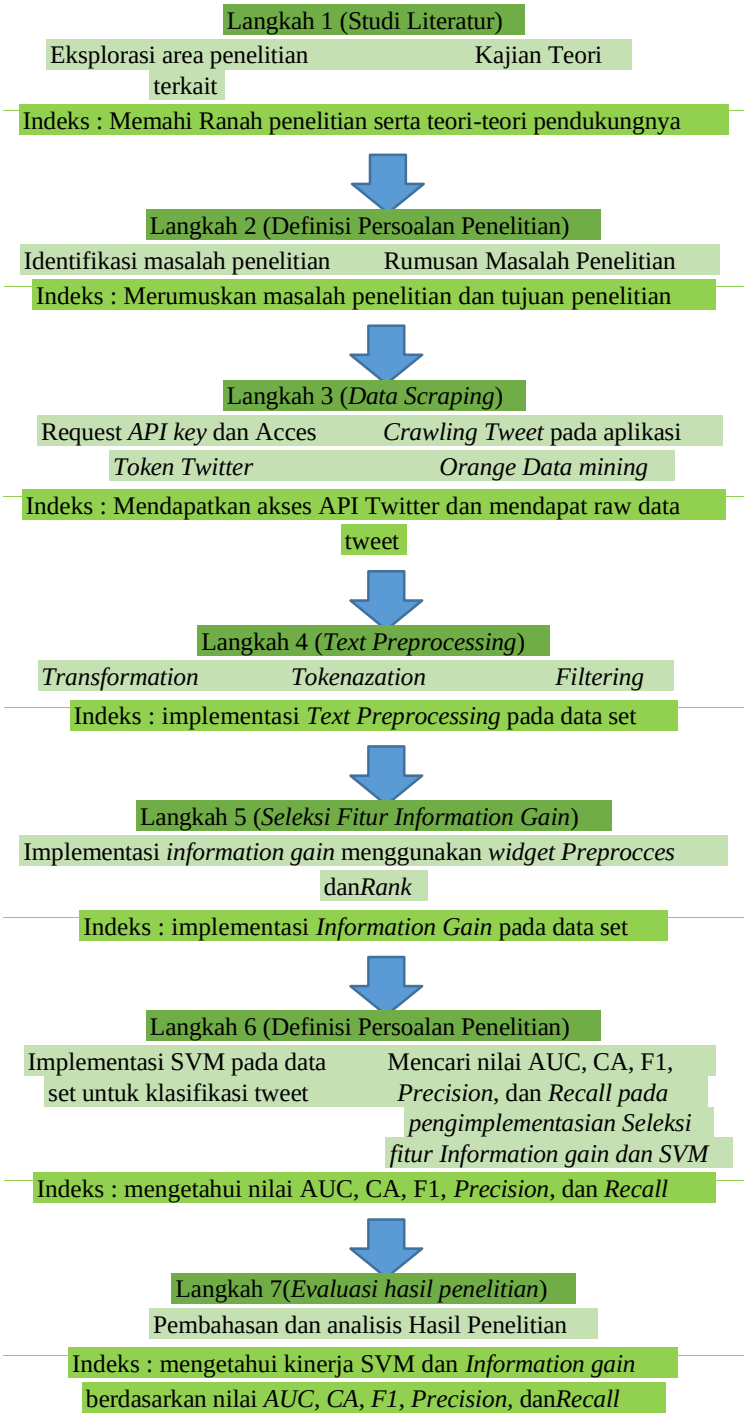
BAB III

METODOLOGI PENELITIAN

3.1. Metode Penelitian

Metodologi penelitian adalah acuan, pedoman ,dan tahapan yang akan diterapkan pada sebuah penelitian untuk mencapai tujuan dari penelitian tersebut. Metodologi penelitian memiliki rencana kerja yang sistematis sehingga hasil yang didapatkan dapat sesuai dengan yang diharapkan. Oleh karena itu diperlukan tahapan yang tersusun. Tahapan pada penelitian ini diawali dengan Studi pustaka dengan mempelajari tentang teori–teori yang menjadi referensi dan pendukung dalam pembuatan dan penulisan hasil penelitian mengenai analisis sentimen pelayanan unit gawat darurat terhadap rumah sakit di Indonesia berdasarkan tweet pada pengguna twitter dengan menggunakan *klasifikasi information gain* dan algoritma *Support Vector Machine* (SVM). Pada tahap pengumpulan data tweet pada pengguna twitter dilakukan dengan menggunakanan twitter API yang telah disediakan oleh twitter untuk tujuan penelitian dan pembelajaran. Pada proses pengambilan data dan analisis sentimen ini akan menggunakan aplikasi*orange data mining* untuk memudahkan proses pengambilan data. Langkah-langkah yang akan dilakukan dalam penelitian ini terdapat pada alur desain penelitian di bawah ini . Secara ringkas alur metode penelitian yang akan dilakukan terdapat pada Gambar 3.1.



Gambar 3. 1 Diagram Alur Penelitian

3.2. Studi Pustaka

Studi pustaka merupakan suatu studi yang digunakan dalam mengumpulkan informasi dan data dengan bantuan berbagai macam sumber dengan melakukan penelaahan terhadap buku, literatur, catatan, serta berbagai laporan yang berkaitan dengan masalah yang ingin dipecahkan. Studi pustaka dilakukan dengan mempelajari tentang teori–teori yang menjadi referensi dan pendukung dalam pembuatan dan penulisan hasil penelitian mengenai analisis sentimen pelayanan unit gawat darurat terhadap rumah sakit di Indonesia berdasarkan tweet pada pengguna twitter dengan menggunakan seleksi fitur *Information Gain* dan algoritma*Support Vector Machine* (SVM).

3.3. Data Scraping

Data *scraping* adalah proses pengambilan data atau esktraksi dari sebuah website, lalu data tersebut umumnya disimpan dalam sebuah format tertentu. Pada penelitian ini data didapat langsung dari web Twitter. *Web scraper* merupakan program dengan yang masuk ke halaman website, mengunduh data, mengekstrak data tergantung pada kebutuhan. Twitter telah menyediakan API *(Application Programming Interface)* nya untuk memudahkan *developer* atau pengguna untuk mengakses informasi semua data tweet yang terdapat pada twitter sesuai dengan yang dibutuhkan.

3.4. Text Preprocessing

*Text preprocessing*adalah suatu proses pengubahan bentuk data yang belum terstruktur menjadi data yang lebih terstruktur sesuai dengan kebutuhan, untuk proses mining lebih lanjut. *Text Preprocessing* adalah tahapan dimana aplikasi

melakukan seleksi data yang akan diproses pada setiap dokumen. Proses *text preprocessing* ini meliputi *case folding, tokenizing, filtering*, dan *stemming.*

3.5. Seleksi Fitur Information Gain

Seleksi fitur adalah teknik untuk memilih fitur penting dan relevan terhadap data dan mengurangi fitur yang tidak relevan atau seleksi fitur juga dapat diartikan sebagai teknik reduksi dimensi yang digunakan untuk memperkecil matriks data dengan memperhatikan informasi kata penting yang perlu diproses. Seleksi fitur berfungsi untuk memilih fitur terbaik dari suatu kumpulan data fitur (Maulida*et al.*, 2016). Banyaknya fitur akan mempengaruhi proses komputasi dan bahkan jika banyak fitur yang tidak relevan digunakan dalam sebuah proses klasifikasi maka dapat juga mempengaruhi hasil akurasinya. Proses pengurangan dimensi data ini juga memungkinkan algoritma klasifikasi untuk bekerja lebih cepat dan efektif serta memungkinkan peningkatan akurasi suatu algoritma.

Information gain merupakan salah satu metode seleksi fitur yang banyak dipakai oleh peneliti untuk menentukan batas dari kepentingan sebuah fitur (Maulana dan Karomi, 2015). *Information Gain* merupakan salah satu algoritma seleksi fitur yang digunakan untuk memilih fitur terbaik. *Information Gain* dimanfaatkan untuk merangking kata-kata penting dari hasil reduksi fitur. Hasil dari proses *Information Gain* adalah kata penting yang bersifat informatif. *Information Gain* merupakan teknik seleksi fitur yang memakai metode scoring untuk nominal ataupun pembobotan atribut kontinu yang didiskretkan menggunakan maksimal entropy. Suatu entropy digunakan untuk mendefinisikan nilai *Information Gain*. Entropy menggambarkan banyaknya informasi yang dibutuhkan untuk

mengkodekan suatu kelas. *Information Gain* dari suatu term diukur dengan menghitung jumlah bit informasi yang diambil dari prediksi kategori dengan ada atau tidaknya term dalam suatu dokumen. Information Gain (IG) ini berguna untuk mengurangi bias agar dapat meningkatkan akurasi dari algoritma (Utami, 2015)

3.6. Text Mining

*Text mining*adalah proses mengeksplorasi dan menganalisis sejumlah besar data teks tidak terstruktur yang dibantu oleh perangkat lunak yang dapat mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lainnya dalam data (Ratniasih, Sudarma dan Gunantara, 2017). Pada penelitian ini text mining digunakan untuk analisis sentimen dan klasifikasi data tweet pada twitter yang mengandung kalimat unit gawat darurat.

3.6.1. Algortima Klasifikasi Support Vector Machine (SVM)

Klasifikasi adalah pembagian atau pengelompokan data atau sesuatu menurut kelas-kelas yang demudian akan mempelajari data latih dengan menggunakan algoritme pengklasifikasian (Pratama, Wihandika dan Ratnawati, 2018). Atau klasifikasi juga dapat diartikan sebagai proses pengelompokkan benda berdasarkan ciri-ciri persamaan dan perbedaan. Klasifikasi merupakan salah satu metode dalam data mining yang bertujuan untuk mendefinisikan kelas dari sebuah objek yang belum diketahui kelasnya. Pada klasifikasi terlebih dahulu akan dilakukan proses training dan testing. Pada proses tersebut akan digunakan dataset yang telah diketahui kelas objeknya.

Masalah klasifikasi adalah bagaimana menentukan suatu objek masuk ke suatu class yang sebenarnya. SVM merupakan salah satu metode yang biasanya

digunakan untuk klasifikasi. Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matnag dan lebih jelas secara matematis. Masalah klasifikasi adalah bagaimana menentukan suatu objek masuk ke suatu class yang sebenarnya.

Hasil yang di dapat dari Implementasi metode *Support Vector Machine* dengan menggunakan Seleksi Fitur *Information gain* adalah nilai AUC, CA, F1, *Precision*, dan *Recall*

Area under the curve (AUC) adalah Matric untuk mengkalkulasi perfoma secara keseluruhan dari model klasifikasi berdasarkan area di bawah kurva ROC.

Accuracy mennggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar.

$$Acuraccy = \frac{True\ Netral + True\ Positif + True\ Negatif}{\text{Jumlah Data}}$$

F-1 Score menggambarkan perbandingan rata-rata precision dan recall yang dibobotkan. Accuracy tepat kita gunakan sebagai acuan performansi algoritma jika dataset kita memiliki jumlah data False Negatif dan False Positif yang sangat mendekati (symmetric). Namun jika jumlahnya tidak mendekati, maka sebaiknya kita menggunakan F1 Score sebagai acuan.

$$F1 = \frac{2 * precision * recall}{precision + recall}$$

Precision menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh mode

$$Precision = \frac{True\ Positif}{True\ Positif + False\ Positif}$$

$$Precision = \frac{Precision\ Negatif + Precision\ Netral + Precision\ Positif}{\text{Jumlah Kelas}}$$

Recall atau sensitivity: menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi.

$$Recall = \frac{True\ Positif}{True\ Positif + False\ Negatif}$$

$$Recall = \frac{Recall\ Negatif + Recall\ Netral + Recall\ Positif}{\text{Jumlah Kelas}}$$