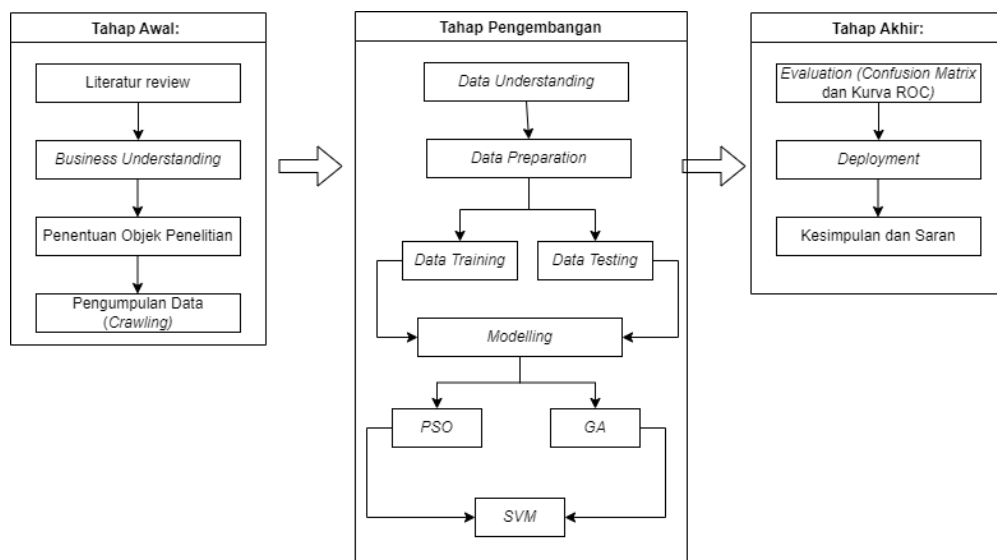


BAB III

METODOLOGI PENELITIAN

Pada penelitian ini menggunakan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Pemrosesan data akan melalui beberapa tahapan seperti berikut :



Gambar 3. 1 Tahapan Penelitian

Tahapan penelitian dibagi kedalam 3 tahap yakni tahap awal yang terdiri dari proses *literature review*, *business understanding*, penentuan objek penelitian, dan pengumpulan data. Kemudian tahap kedua yakni tahap pengembangan yang terdiri dari proses *data understanding*, *data preparation*, *data training*, kemudian hasil dari proses *data training* digunakan untuk memproses *data testing*, selanjutnya tahap *modelling* menggunakan algoritma SVM serta *feature selection* PSO dan GA. Tahap terakhir yakni meliputi proses *Evaluation* dalam bentuk

confusion matrix dan kurva ROC, kemudian tahap *deployment* dan tahap terakhir pengambilan simpulan dan saran. Penjelasan mengenai proses pada tahapan penelitian sebagai berikut:

3.1 Tahap Awal

3.1.1 Literatur *Review*

Pada tahap literatur *review* dilakukan pencarian terkait penelitian sebelumnya yang berhubungan dengan penelitian ini. Studi literatur diperoleh dari jurnal terkait algoritma dan metode yang digunakan dalam penelitian. Sumber literatur lain yang digunakan diantaranya buku, skripsi terdahulu, dan sumber Pustaka dari internet.

3.1.2 *Business Understanding*

Pada tahap ini akan diidentifikasi kebutuhan yang diperlukan dalam lingkup penelitian untuk menentukan tujuan. Kemudian tujuan dan batasan akan diterjemahkan untuk menentukan permasalahan yang akan diangkat. *Business Understanding* merupakan langkah awal dalam melakukan proses penelitian yang bertujuan untuk mendefinisikan permasalahan yang terjadi supaya masalah tersebut menjadi lebih terukur sebagai Langkah awal penelitian. Setelah mengidentifikasi masalah Langkah selanjutnya yakni mendefinisikan masalah penelitian.

3.1.3 Penentuan Objek Penelitian

Penentuan objek penelitian dilakukan dengan proses analisis untuk menentukan suatu kasus yang akan dijalankan sebagai objek penelitian. Setelah

dilakukan proses analisis pada domain penelitian maka akan didapatkan suatu objek penelitian.

3.1.4 Pengumpulan Data

Data yang diperlukan dalam penelitian merupakan data tweet pada sosial media Twitter terkait dengan Pemekaran Papua untuk dijadikan sampel dalam membandingkan kinerja algoritma *Support Vector Machine* (SVM) dengan menerapkan *feature selection Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA) untuk memperoleh hasil akurasi terbaik.

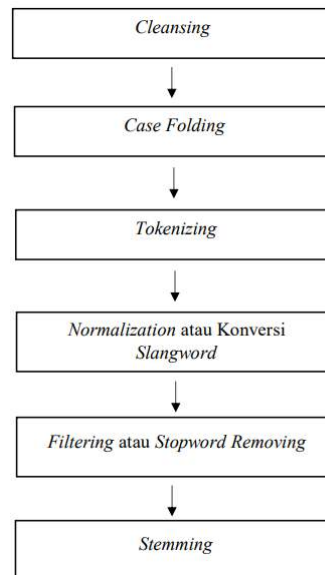
3.2 Tahap Pengembangan

3.2.1 Data Understanding

Data *Understanding* digunakan untuk memeriksa data, sehingga dapat mengidentifikasi masalah dalam data. Tahap ini bertujuan untuk mengumpulkan, mengidentifikasi, dan memahami data yang ada.

3.2.2 Data Preparation

Data yang sudah terkumpul kemudian akan diberikan sentimen. Setelah diberi sentimen pada data selanjutnya akan dilakukan tahap *pre-processing*. Tahap *pre-processing* dilakukan untuk memperbaiki data tidak terstruktur atau untuk menghilangkan *noise* pada data menjadi data terstruktur untuk mempermudah proses analisis. Dengan kata lain tahap *pre-processing* diperlukan untuk menemukan data yang diperlukan dan menghilangkan data yang tidak diperlukan. Tahapan yang akan dilakukan pada tahap *pre-processing* diantaranya :



Gambar 3. 2 Tahapan Pre-Processing

3.2.3 Modelling

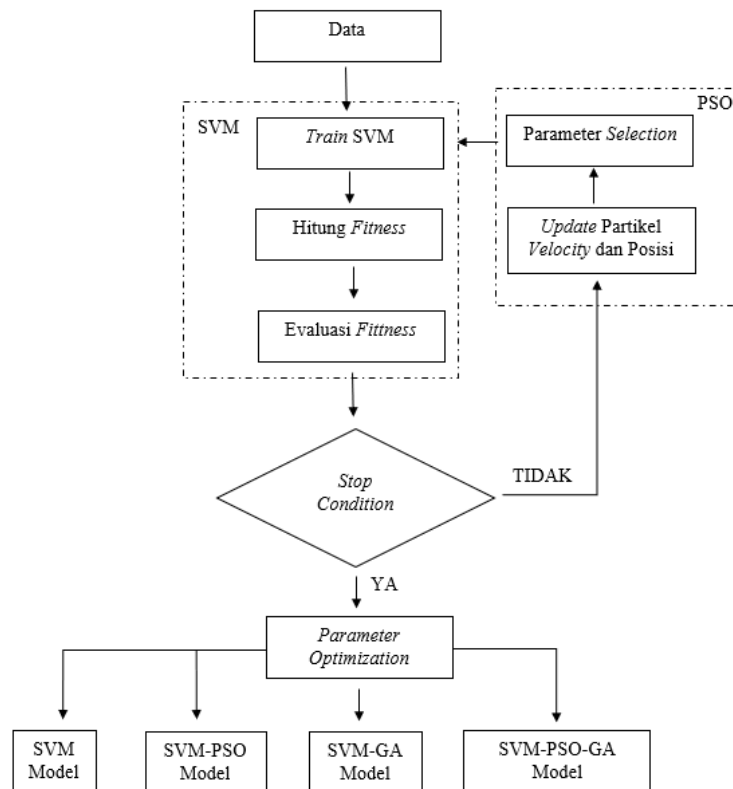
Setelah melewati tahap *pre-processing* selanjutnya akan dilakukan pembobotan atau *term weighting* berdasarkan fitur TF-IDF. Kata yang diambil sebagai penentuan data yang termasuk dalam kategori positif seperti dukung, mendukung, pemerataan pembangunan, setuju, terang, kesejahteraan; Kategori negatif seperti tolak, menolak, menghambat, demo, aksi. Hasil yang akan didapat dari pengolahan data teks berupa bobot nilai yang berada dalam masing-masing data tweet untuk kemudian dilakukan perbandingan setiap probabilitas nilai setiap data berdasarkan pembobotan TF-IDF.

Hasil yang didapatkan dari probabilitas dengan atribut positif dan atribut negatif akan dibandingkan untuk mengetahui atribut yang memiliki probabilitas paling besar. Apabila nilai probabilitas opini negatif lebih besar daripada opini

positif maka akan ditetapkan bahwa dokumen tersebut merupakan opini negatif, begitupun sebaliknya apabila nilai probabilitas opini positif lebih besar daripada opini negatif maka akan ditetapkan bahwa dokumen tersebut merupakan opini positif. Namun apabila probabilitas opini negatif dan opini positif memiliki probabilitas nilai yang sama maka ditetapkan dokumen tersebut termasuk kategori opini netral.

Data tweet akan dipecah menjadi data *training* dan data *testing*. Pada tahap ini akan menggunakan 10 *k-fold cross validation*. Dimana 10 *k-fold cross validation* termasuk metode yang digunakan untuk menghilangkan *noise* atau bias kata untuk menaikkan nilai akurasi. Pada 10 *K-fold Cross-validation* dilakukan pemisahan data secara acak ke dalam 10 bagian untuk dilakukan proses klasifikasi data.

Data *training* digunakan untuk melatih algoritma sedangkan data *testing* digunakan untuk menentukan kinerja algoritma yang telah dilatih sebelumnya. Hal yang dilakukan pertama dalam proses analisis menggunakan SVM adalah mengubah teks menjadi data vektor. Karena kesulitan dalam menentukan parameter yang kompatibel dengan SVM, maka pada penelitian ini menerapkan dua teknik optimasi yakni PSO dan GA untuk menetapkan parameter SVM untuk pengoptimalan. Pada penelitian ini nilai default setiap parameter SVM dan SVM-PSO adalah 0.



Gambar 3. 3 Model SVM-PSO

3.3 Tahap Akhir

3.3.1 Evaluation

Proses evaluasi dilakukan dengan tujuan untuk menguji performa model dari algoritma *Support Vector Machine* (SVM). Proses evaluasi akan dilakukan menggunakan *Confusion Matrix* untuk mendapatkan nilai *accuracy*, *precision*, *recall*, dan *f-score* dengan hasil data dalam bentuk kurva ROC untuk mengukur nilai AUC.

3.3.2 *Deployment*

Tahapan terakhir yang akan dilalui dari metode CRISP-DM yakni tahap penyebaran atau *Deployment* yakni mempresentasikan hasil dari model yang telah digunakan pada proses data mining. Penyajian hasil proses yang meliputi pengetahuan yang didapat selama proses sehingga dapat dipahami oleh pengguna. Kemudian pembuatan laporan hasil penelitian.

3.3.3 Kesimpulan dan Saran

Tindakan yang akan dilakukan pada tahap ini yakni pemberian kesimpulan terhadap proses penelitian yang telah dilakukan serta memberikan saran untuk penelitian yang akan dilakukan selanjutnya oleh peneliti lain.