

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Kesimpulan yang diperoleh sebagai berikut :

1. Membangun model *machine learning* dari dataset sentimen masyarakat Jakarta terhadap isu tenggelamnya Jakarta yang sudah memiliki label positif dan negatif menggunakan algoritma *support vector machine* dapat dilakukan yang pertama kali adalah melakukan praproses pada data, seperti memperbaiki data yang tidak seimbang menggunakan *undersampling* dan *oversampling*, melakukan *casefolding* agar data tetap konsisten, penghapusan simbol, melakukan *tokenizing*, penghapusan *stopword*, dan mengubah kata imbuhan menjadi kata dasar menggunakan *stemming*. Setelah selesai dilakukan pra-proses, lalu dilakukan tahap TF-IDF, kemudian data dibagi menjadi data latih dan data uji untuk pembuatan model SVM. Pembagian data latih dan data uji ini menggunakan 2 jenis perbandingan data yaitu perbandingan pertama menjadi 80% data latih dan 20% data uji dan perbandingan kedua menjadi 70% data latih dan 30% data uji, lalu diklasifikasi menggunakan support vector machine dengan parameter *cost* dan kernel yang berbeda untuk mencari hasil evaluasi yang terbaik. Setelah diklasifikasi model yang telah dibuat kemudian dilakukan evaluasi menggunakan *confusion matrix* untuk melihat nilai hasil akurasi, *precision*, *recall*, *specificity* dan *F1-Score* dari dataset original, dataset *undersampling* dan dataset *oversampling* untuk mencari nilai rata-rata hasil evaluasi yang terbaik diantara ketiga jenis dataset.
2. Dari penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma *support vector machine* pada dataset sentimen masyarakat Jakarta mengenai kinerja pemerintah Jakarta pada isu tenggelamnya Jakarta menghasilkan hasil evaluasi yang bagus yaitu terdapat pada perbandingan 80% data latih dan 20% data uji menggunakan dataset *oversampling* dengan nilai rata-rata akurasi sebesar 0,9359, *precision* sebesar 0,9676, *recall* sebesar 0,9003, *specificity* sebesar 0,9703 dan *f1-score* sebesar 0,9326. Sementara dataset original menghasilkan nilai rata-rata akurasi terbaik pada perbandingan 80:20 dengan akurasi sebesar 0,8147, *precision* sebesar 0,8532, *recall* sebesar 0,8333, *specificity* sebesar 0,7883 dan *f1-score* sebesar 0,8415. Untuk dataset *undersampling* menghasilkan nilai rata-rata akurasi terbaik pada perbandingan 70:30 dengan akurasi sebesar 0,7534, *precision* sebesar 0,8162, *recall* sebesar 0,6562, *specificity* sebesar 0,8506 dan *f1-score* sebesar 0,7269.

## 5.2 Saran

Adapun saran yang dapat diberikan:

1. Menambah jumlah data sehingga data menjadi lebih banyak.
2. Parameter yang lebih variatif, sehingga memungkinkan untuk mencapai nilai akurasi terbaik.
3. Menggunakan teknik *smote*, *adasyn* dan teknik yang lain dalam memperbaiki data yang tidak seimbang.