

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berdasarkan penelitian kombinasi algoritma *Support Vector Machine* dan *Lexicon-based* untuk analisis sentimen pada Peraturan Pemerintah No 25 Tentang Tabungan Perumahan Rakyat Tahun 2020, diperoleh kesimpulan bahwa:

1. Kesulitan pada proses pelabelan dataset manual yang membutuhkan waktu yang banyak, bisa diselesaikan dengan *lexicon-based*. *Lexicon-based* adalah sebuah metode pelabelan berdasarkan kamus *lexicon* yang telah dibuat. Kamus ini berisi kumpulan kosakata atau istilah berupa sentimen atau opini yang memiliki bobot positif dan negatif. Proses pelabelan *lexicon-based* berhasil memberikan label terhadap dataset sebanyak 647 *tweet* dengan rincian 424 *tweet* berlabel positif, 175 *tweet* berlabel negatif, dan 48 data *tweet* berlabel netral. Hasil dari proses *lexicon-based* digunakan untuk mentransfer pembelajaran kepada SVM dalam bentuk pelabelan data dan sebagai bentuk kombinasi *lexicon based* dan algoritma SVM.
2. Pengujian analisis sentimen tentang Peraturan Pemerintah No 25 Tentang Tabungan Perumahan Rakyat Tahun 2020 dengan 3 label sentimen, menggunakan empat fungsi *kernel* pada SVM untuk dilihat performansi akurasi terbaik. Fungsi *kernel* yang digunakan adalah *kernel linear*, *kernel rbf*, *kernel polynomial*, dan *kernel sigmoid*. Nilai *accuracy* terbaik didapat dari fungsi *kernel linear* dengan nilai *hyperparameter C* sebesar 10, nilai *max iteration* sebesar 100 dengan nilai *accuracy* sebesar 83.08%.

## 5.2. Saran

Berdasarkan penelitian yang telah dilakukan, ada beberapa saran yang bisa diberikan untuk penelitian selanjutnya, yaitu:

1. Menambahkan *sentiment lexicon* seperti *NRC Emotion Lexicon*. Keuntungan dari *NRC Emotion Lexicon* merupakan *lexicon* yang dilatih mengenai *tweet* dengan memanfaatkan keterhubungan tiap kata dengan delapan emosi dasar (kemarahan, ketakutan, antisipasi, kepercayaan, kejutan, kesedihan, kegembiraan, dan *disgust*) pada *lexicon*. Hal ini bisa dilakukan sebagai cara untuk memperbaiki metode pelabelan *lexicon-based* dengan ditambahkan berdasarkan hubungan pada delapan emosi dasar.
2. Menggunakan metode *resampling* jika mengalami perbedaan distribusi label atau kelas data yang signifikan atau *unbalanced dataset*. Dikarenakan, *unbalanced dataset* memiliki potensi untuk mengganggu kinerja model *machine learning*.
3. Penelitian selanjutnya, bisa dikembangkan untuk dilakukan beberapa skenario atau percobaan pengujian. Seperti, perbandingan komposisi data training dan data testing, menggunakan tahapan *preprocessing lemmatization*, dsb.