

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan pada penelitian yang telah dilaksanakan, dapat ditarik kesimpulan sebagai berikut:

1. Data yang dimanfaatkan termasuk ke dalam data tweet dari media sosial twitter mengenai Pekan Olahraga Nasional XX yang di *crawling* dari tanggal 01 Agustus 2021 hingga 31 Desember 2021 sebanyak 1000 ulasan. Setelah itu data di lakukan pembersihan dengan menghilangkan data *duplicate* sehingga memperoleh hasil 218 data. Data dilaksanakan pelabelan dengan cara manual dan didapatkan jumlah kelas masing-masing sebanyak 199 positif dan 19 negatif. Kemudian data dilaksanakan pembersihan terlebih dahulu di praproses sebelum dilaksanakan pembobotan, kemudian data yang sudah bersih diberikan bobot setiap kata dengan *Term Frequency-Invers Document Frequency* (TF-IDF) yang nantinya akan dijadikan sebagai fitur. Sebelum fitur tersebut akan dilaksanakan pembagian data menjadi data latih dan data uji, dilaksanakan tahapan penyeimbangan data agar data berlabel positif dan negatif seimbang untuk diolah dengan metode *Synthetic Minority Oversampling Technique* (SMOTE). Setelah fitur seimbang, lalu fitur tersebut dibagi dua menjadi data *training* dan data *testing* untuk membentuk suatu model dengan memanfaatkan metode *Naïve Bayes*.
2. Performa pengujian model yang terbentuk dengan metode *Naïve Bayes* dilaksanakan dengan membandingkan hasil dari model tersebut dengan data uji sebanyak 20% dari keseluruhan data. Selanjutnya dilaksanakan evaluasi tersebut didapatkan hasil akurasi yang besarnya yakni 99%, dengan nilai presisi 100%, *recall* 98%, dan *specitifity* 100%.

#### 5.2 Saran

Berdasarkan hasil analisis dan pembahasan pada penelitian yang telah dilaksanakan, dapat diberikan saran sebagai berikut:

1. Diharapkan untuk penelitian berikutnya untuk meningkatkan pada proses pre-processing dan pemodelan data yang lebih baik sehingga bisa menganalisis dengan baik dan evaluasi model yang sangat baik.
2. Pada tahap normalisasi data diharapkan untuk menambahkan lagi kosa kata pada kamus untuk menghindari adanya singkatan kata dan *slang word* sehingga makna kata dapat dimengerti dan formal
3. Label yang dimanfaatkan tidak hanya positif dan negatif saja, tetapi menambahkan label netral.
4. Penelitian selanjutnya bisa memanfaatkan algoritma klasifikasi lainnya seperti *Support Vector Machine*, *K-Nearest Neighbor* dan lain-lain sebagai perbandingan untuk performa model.