

BAB 5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian dari analisis sentimen pengaruh Covid-19 terhadap kehidupan bermasyarakat pada media sosial *Twitter* menggunakan K-Nearest Neighbor, maka dapat diambil kesimpulan sebagai berikut:

1. Berdasarkan data yang diambil dari *Twitter* pada tanggal 5 Desember 2020 sampai 12 Desember 2020, didapatkan sebanyak 420 dokumen dari Jakarta, dan 685 dokumen dari Surabaya. Dari data yang didapatkan, dilakukan pelabelan yang dilakukan oleh 3 orang *annotator* yang mendapatkan hasil bahwa pada data Jakarta didapatkan 268 *tweet* dinyatakan positif, dimana mewakili 63.80% dokumen, dibandingkan dengan 152 *tweet* dinyatakan negatif, yang mewakili 36.20% dokumen. Sementara itu, pada data Surabaya didapatkan 398 *tweet* positif, yang mewakili 58.10% dokumen, dengan 287 *tweet* negatif, yang mewakili 41.90% dokumen. Dari kedua data Jakarta dan Surabaya dapat dilihat bahwa dengan adanya Covid-19, penggunaan media sosial *Twitter* dalam membahas topik Covid-19 lebih banyak berisikan hal-hal positif dibandingkan dengan hal-hal negatif.
2. Dalam melakukan klasifikasi data *tweet* menggunakan K-Nearest Neighbor, dilakukan *text preprocessing*, yang terbagi atas *case folding*, *stemming*, *filtering*, *tokenization*, dan normalisasi Bahasa. Setelah itu, dilakukan *undersampling* untuk menyeimbangkan *sentiment* positif, dan negatif. Setelah itu, dilakukan pembobotan dokumen menggunakan TF-IDF untuk masing-masing data Jakarta, dan Surabaya untuk diteruskan menuju penggunaan KNN. Dalam percobaan menggunakan KNN, dilakukan 3 kali percobaan yang menggunakan 70%/75%/80% data *train*, 30%/25%/20% data *test*, dengan *random state* 42 yang akan diujikan dengan nilai K dari K=1,

sampai $K=20$. Pada data Jakarta, didapatkan akurasi terbaik pada percobaan 2 dengan $K=9$, yaitu 78.85%. Untuk data Surabaya, didapatkan akurasi terbaik pada percobaan 1 dengan $K=(11,17,19,20)$, yaitu 76.47%.

5.2. Saran

Berdasarkan penelitian ini, maka didapatkan beberapa masukan untuk melakukan penelitian selanjutnya, yaitu:

1. Menggunakan teknik normalisasi Bahasa selain dari *word by word* menggunakan *colloquial dictionary*, yaitu normalisasi Bahasa per kalimat *tweet*.
2. Menggunakan teknik pembagian data selain train test split, seperti *stratified K-fold* dengan menguji tiap *fold*nya.
3. Dalam menggunakan klasifikasi KNN, bisa menggunakan metode perhitungan selain *Euclidian Distance*, seperti *Manhattan Distance*, *Minkowski*.