

BAB V

KESIMPULAN

5.1 Kesimpulan

Penelitian ini melakukan pengujian model menggunakan Naïve Bayes Classifier (NBC) yang memakai proses Part of Speech (POS) Filtering dan tidak memakai POS Filtering. Pengujian dilakukan dengan data tweet yang telah dikumpulkan melalui proses pengunduhan twitter menggunakan twitter scraper. Setelah data terkumpul, data yang akan digunakan untuk proses training akan diberi label sentimen positive, negative, ataupun neutral. Data training ini nantinya akan diolah terlebih dahulu melalui serangkaian proses. Terdapat langkah Preprocessing, Pembentukan Feature List, Vektor Feature, Perhitungan TF-IDF dan Cosine Similarity. Adapun proses Preprocessing yang dilakukan yaitu :

- a. Tweet Cleaning
- b. Tokenize
- c. Slang Replacement
- d. Remove Stopword
- e. POS Tagging
- f. POS Filtering
- g. Stemming
- h. Pembentukan Feature List

Setelah dilakukan preprocessing, Masing-masing tweet akan diberi bobot menggunakan metode TF-IDF terlebih dahulu sebelum memasuki perhitungan Cosine Similarity. Cosine Similarity digunakan untuk memberi bobot kedekatan terhadap feature list positive, negative, dan neutral. Kemudian bobot kedekatan ini akan diklasifikasi dengan algoritma Naïve Bayes Classifier dan akan dilakukan proses evaluasi terhadap hasil pelatihan maupun prediksi yang dilakukan oleh algoritma Naïve Bayes Classifier.

Berikut adalah proses evaluasi yang dilakukan dalam penelitian :

1. Proses pengujian pertama yaitu melakukan perbandingan jumlah fitur yang diperoleh dari tahap feature selection yang dilakukan menggunakan POS Filtering dan tanpa menggunakan POS Filtering. Jumlah fitur yang dihasilkan oleh POS Filter yaitu 911 kata dengan jumlah fitur positive 199 kata, negative 444 kata, dan neutral 268 kata. Sedangkan jumlah fitur yang dihasilkan tanpa proses POS Filter 1.297 kata dengan jumlah fitur positive 292 kata, negative 588 kata, dan neutral 417 kata. Dengan demikian dapat disimpulkan bahwa proses POS Filtering berhasil menyaring feature yang akan dihasilkan berdasarkan tag dari masing-masing kata yang diproses.
2. Skenario pengujian kedua yaitu melakukan evaluasi dari hasil pelatihan klasifikasi sentimen Naïve Bayes Classifier yang menggunakan POS Filtering, diperoleh akurasi model sebesar 89%. Sedangkan akurasi model tanpa menggunakan POS

Filtering memperoleh akurasi model sebesar 95%. Dengan demikian dapat disimpulkan bahwa jumlah kata pada feature list yang digunakan pada proses pembobotan Cosine Similarity sangat berpengaruh terhadap proses pelatihan algoritma klasifikasi yang dilakukan oleh kedua model.

3. Skenario yang ketiga yaitu melakukan evaluasi terhadap data prediksi. Masing-masing model yang dihasilkan oleh Naïve Bayes Classifier dengan proses POS Filtering maupun tanpa proses POS Filtering akan dievaluasi hasil prediksi sentimen terhadap data prediksi. Data prediksi merupakan data yang sudah dipisahkan dari proses pelatihan model. Hasil prediksi yang dilakukan oleh model dengan POS Filtering mendapatkan akurasi prediksi sebesar 87%. Berbeda dengan model tanpa POS Filtering yang menghasilkan akurasi prediksi lebih besar yaitu 90%. Dengan demikian dapat disimpulkan bahwa jumlah kata pada feature list yang digunakan pada proses pembobotan Cosine Similarity sangat berpengaruh terhadap presentase akurasi proses prediksi yang dilakukan oleh algoritma NBC.

5.2 Saran

Pada penelitian ini, terdapat beberapa keterbatasan dan kekurangan. Kekurangan dan keterbatasan ini bisa dijadikan acuan dan pertimbangan

untuk penelitian selanjutnya. Adapun saran yang dihasilkan dalam penelitian ini yaitu sebagai berikut:

1. Jumlah data training yang dipakai dalam proses pelatihan diusahakan tidak ada tweet yang sama.
2. Jumlah kata yang terdapat pada feature list sebisa mungkin dibuat rata ataupun sama karena dapat mempengaruhi akurasi dari model Naïve Bayes Classifier.
3. Korpus untuk melakukan POS Tagging dalam penelitian ini merupakan korpus yang terbentuk dari kalimat-kalimat formal. Sedangkan tweet tidak selalu terdiri dari kata baku. Perlu adanya korpus POS Tagger tersendiri yang berisikan pola-pola kalimat tidak baku.
4. Pembuatan kamus untuk menerjemahkan kalimat-kalimat tidak baku yang biasa digunakan pengguna Twitter.