

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan dari penelitian Implementasi Algoritma K-Means Dan *K-Nearest Neighbor* Untuk *Topic detection and tracking* Pada *Microblog* Twitter adalah sebagai berikut.

1. Algoritma K-Means dapat membantu mendeteksi adanya topik pada data twitter dengan cara mengelompokkan *tweet* berdasarkan nilai *cosine similarity* terbesar antara *tweet* dengan *centroid* setiap *cluster*. K-Means membutuhkan parameter jumlah *cluster* (k). Nilai k sangat berpengaruh terhadap hasil *cluster*. Setiap *cluster* yang dihasilkan dari K-Means ini merepresentasikan sebuah topik yang dideteksi, sehingga dapat diartikan bahwa satu *cluster* mewakili satu topik. Sementara itu k-NN sebagai algoritma yang dipakai untuk proses klasifikasi dapat mengklasifikasikan *tweet* berdasarkan nilai *cosine similarity* terbesar antara *tweet* yang akan diklasifikasikan dengan *tweet* yang telah melalui proses *clustering*. Algoritma ini membutuhkan parameter berupa nilai k yang mewakili jumlah *tweet* terdekat yang akan dibandingkan. Pada penelitian ini digunakan nilai *threshold* sebagai batas apakah *tweet* tersebut diklasifikasikan ke dalam topik yang baru atau topik yang sudah ada. Nilai *threshold* berkisar antara 0 sampai 1. Nilai *threshold* dibandingkan dengan nilai *cosine similarity*, jika nilai *cosine similarity* lebih kecil daripada nilai *threshold* maka *tweet* tersebut diklasifikasikan sebagai topik baru, sedangkan jika nilai *cosine similarity* lebih besar daripada nilai *threshold* maka *tweet* tersebut diklasifikasikan ke dalam topik yang sudah ada.
2. Hasil dari aplikasi ini mendapatkan respon baik dari responden. Secara keseluruhan responden menilai bahwa aplikasi ini bermanfaat untuk

mendapatkan informasi tentang Bandung melalui twitter namun dengan beberapa perbaikan agar hasil yang didapatkan lebih akurat.

5.2 Saran

Berikut merupakan saran-saran pada penelitian ini untuk pengembangan lebih lanjut:

1. Perlu dilakukan *preprocessing* terhadap *tweet* yang berisi bahasa daerah dengan cara menambahkan kamus kata bahasa daerah agar data yang dihasilkan lebih berkualitas.
2. Perlu dilakukan perhitungan *cosine similarity* berdasarkan *tweet*, waktu dan lokasi agar topik yang dihasilkan kualitasnya lebih baik.
3. Perlu ditambahkan nilai *threshold* ketika melakukan *clustering* dengan K-Means agar jumlah *cluster* yang dihasilkan fleksibel.
4. Perlu dilakukan penelitian untuk kasus yang sama dengan menggunakan algoritma *clustering* yang lain kemudian membandingkan hasilnya dengan algoritma K-Means.
5. Perlu dilakukan penelitian untuk kasus yang sama dengan menggunakan algoritma klasifikasi yang lain kemudian membandingkan hasilnya dengan algoritma k-NN.
6. Perlu dilakukan penelitian lain untuk kasus yang sama dengan teknik ekstraksi informasi yang lain secara detail sehingga hasil dari *keyword extraction* dapat lebih baik.