BAB V
KESIMPULAN DAN SARAN

1.1. Kesimpulan

Kesimpulan dari penelitian Clustering Keluhan Kota Bandung dengan Memanfaatkan Data Twitter Menggunakan Algoritma K-Means adalah sebagai berikut.

1. Data Twitter dapat digunakan untuk mengumpulkan keluhan masyarakat Kota Bandung, namun tidak dapat langsung digunakan karena data *tweet* mengandung banyak *noise*, spam, serta duplikasi. Diperlukan praproses sebelum masuk ke dalam proses klasifikasi maupun *clustering*. Dengan melakukan proses klasifikasi menggunakan algoritma kNN terlebih dahulu, data dapat dikelompokkan ke dalam *tweet* keluhan dan bukan keluhan sehingga lebih mudah diproses di tahapan *clustering*.
2. Proses klasifikasi menggunakan algoritma kNN mendapatkan akurasi tertinggi sebesar 76,6% dengan nilai k=1 dan rata-rata akurasi sebesar 75,06% dengan nilai k=1 hingga k=10.
3. Pada proses *clustering* menggunakan k-Means, *tweet* diproses dalam rentang waktu per bulan dan per minggu untuk melihat keluhan dalam rentang waktu tersebut. Selain itu, dilakukan pengelompokkan *tweet* keluhan bukan macet dan macet.
4. Hasil proses *clustering* untuk rentang waktu per minggu, keluhan bukan macet menghasilkan *purity* tertinggi sebesar 0,8064 pada k=6. Sedangkan untuk keluhan macet menghasilkan *purity* tertinggi sebesar 0,8464 pada k=13. Sementara itu, untuk rentang waktu per bulan, keluhan bukan macet menghasilkan *purity* tertinggi sebesar 0,6422 pada k=13. Sedangkan untuk keluhan macet menghasilkan *purity* tertinggi sebesar 0,6089 pada k=29. Dari hasil *clustering* tersebut, penggunaan algoritma k-Means lebih efektif dilakukan pada rentang waktu per minggu.

Tommy Putra Pratama Gunawan, 2017
CLUSTERING KELUHAN KOTA BANDUNG DENGAN MEMANFAATKAN DATA TWITTER MENGGUNAKAN ALGORITMA K-MEANS
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

1.2. Saran

Saran-saran yang diberikan pada penelitian ini untuk pengembangan lebih lanjut adalah sebagai berikut.

1. Perlu menambahkan kata kunci yang lebih banyak untuk mendapatkan variasi *tweet* keluhan yang lebih banyak sehingga mendapatkan data yang lebih banyak lagi.
2. Perlu dilakukan praproses teks yang lebih baik agar data yang digunakan lebih berkualitas dan meminimalkan *noise*. Cara-cara yang dapat dilakukan adalah seperti menambahkan kamus kata baku, menambahkan daftar *stopword* dan sinonim.
3. Perlu dilakukan penelitian untuk kasus yang sama dengan menggunakan algoritma klasifikasi dan *clustering* yang berbeda untuk membandingkan akurasi.
4. Perlu perubahan pembentukan *cluster* standar yang lebih baik agar menghasilkan nilai *purity* yang lebih baik.

Tommy Putra Pratama Gunawan, 2017
CLUSTERING KELUHAN KOTA BANDUNG DENGAN MEMANFAATKAN DATA TWITTER MENGGUNAKAN ALGORITMA K-MEANS
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Tommy Putra Pratama Gunawan, 2017
CLUSTERING KELUHAN KOTA BANDUNG DENGAN MEMANFAATKAN DATA TWITTER MENGGUNAKAN ALGORITMA K-MEANS
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu