

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dapat disimpulkan beberapa hal sebagai berikut :

1. Algoritma K-Means dapat diterapkan dalam mengelompokkan berita dengan menggunakan representasi dokumen model ruang vektor.
2. Algoritma K-Means melakukan inisialisasi *centroid* secara acak, hal ini membuat hasil akhir dari *clustering* adalah partisi suboptimal. solusi dari permasalahan ini adalah melakukan inisialisasi jumlah *cluster* berulang kali untuk mendapatkan hasil *clustering* yang terbaik. Sehingga membuat algoritma ini sangat tergantung pada inisialisasi jumlah *cluster*.
3. Hasil Eksperimen memperlihatkan bahwa hasil *cluster* terbaik dibangun dengan penambahan proses klasifikasi sebelum proses clustering dengan nilai *Purity*=0,53 dibandingkan dengan tidak diklasifikasi sebelumnya yang memiliki nilai *purity*=0,32. Sehingga dataset masukan yang diberikan adalah dokumen berita yang telah diklasifikasi terlebih dahulu karena memiliki tingkat kemurnian *cluster* yang lebih tinggi dibandingkan dengan dokumen berita yang tidak diklasifikasi terlebih dahulu.

5.2 Saran

Untuk mengembangkan penelitian ini, dapat dipertimbangkan berapa hal sebagai berikut:

1. Evaluasi penggunaan algoritma *clustering* lain yang dapat digunakan untuk melakukan proses *clustering*. Pada Penelitian ini, hanya menggunakan algoritma K-Means.
2. Pemrosesan artikel yang berasal dari sumber multi bahasa. Sehingga tidak hanya terbatas pada berita berbahasa Indonesia. Perlu dilakukan deteksi bahasa secara otomatis dari sebuah dokumen. Selain itu, perlu diperhatikan lagi tahapan *preprocessing text* untuk setiap bahasa yang berbeda.
3. Penanganan pada *outlier* sangatlah penting, dengan mengkaji pembuatan *outlier detection*, sehingga berita yang berisi peristiwa yang baru terjadi tidak akan masuk dianggap sebagai *outlier*.
4. Pengkajian terhadap algoritma stemmer bahasa Indonesia.

