

## BAB 6

### KESIMPULAN DAN SARAN

Pada bab ini akan dibahas kesimpulan dari awal hingga akhir penelitian dan saran untuk penelitian selanjutnya

#### 6.1 Kesimpulan

Kesimpulan diperoleh setelah melakukan beberapa langkah pengerjaan. Berdasarkan hasil pengerjaan didapat kesimpulan bahwa:

1. Untuk melakukan implementasi *clustering* menggunakan dataset yang diuji pada skripsi ini perlu dilakukan *preprocessing* terlebih dahulu, dimana setiap objek masukan berupa dokumen diubah ke dalam representasi vektor.
2. Pada penerapan *clustering* menggunakan algoritma PSO, partikel direpresentasikan sebagai matriks kumpulan *centroid* untuk setiap partikelnya.
3. Berdasarkan *dataset* dan kombinasi parameter yang diuji nilai *purity* yang dihasilkan PSO lebih tinggi 18.2% dari *K-means*. Namun waktu yang di perlukan untuk menyelesaikan program pada PSO lebih lama 81% dari *K-means*.
4. Berdasarkan dokumen berbasis teks yang diuji pada skripsi ini, dapat digunakan fungsi *fitness* seperti pada formula 2.13. Dimana untuk menghitung jarak antar *cluster* dihitung dahulu nilai kemiripan menggunakan rumus *cosine similarity* dan kemudian dijumlahkan seluruh hasil perhitungan kemiripan untuk setiap objek yang ada dalam *cluster* dan untuk setiap *cluster* yang ada dalam partikel.

#### 6.2 Saran

Bagian ini akan membahas saran dari penulis untuk pengembangan penelitian lebih lanjut. Berikut merupakan saran dari penulis:

- Pada penelitian ini, walaupun pada implementasi PSO *cluster* yang dihasilkan berhasil memiliki nilai *purity* yang cukup tinggi (0.94), namun hasil *cluster* belum sepenuhnya sesuai dengan hasil pengelompokan manual. Dari kejadian ini dapat disimpulkan bahwa, walaupun dokumen memiliki jarak vektor yang dekat, namun bisa saja dokumen tersebut memiliki topik yang berjauhan. Hal ini dapat dipengaruhi oleh cara implementasi *cosine distance* yang kurang tepat dalam menghitung jarak antar vektor. Berdasarkan hasil dari pengujian dari skripsi ini, penulis menyarankan untuk menentukan dahulu kumpulan *term/fitur/kata* yang mana dapat digunakan sebagai pembeda antar *cluster* dan kemudian perhitungan *cosine distance* dilakukan hanya menggunakan kumpulan kata tersebut.
- Untuk meningkatkan hasil *clustering*, pada penelitian lebih lanjut dapat dilakukan *filter stopwords* dan juga *stemming* pada penerapan *preprocessing*.

## DAFTAR REFERENSI

- [1] Kadhim, A. I. (2018) An evaluation of preprocessing techniques for text classification. *International Journal of Computer Science and Information Security*, **16**, 22–32.
- [2] Knuth, D. E. (1998) The art of computer programming. Bagian dari Harlow (ed.), *The Art Of Computer Programming*. Addison Wesley Longman, Boston.
- [3] Sailaja, D., Kishore, M., Yothi, B., dan Prasad, N. (2015) An overview of pre-processing text clustering methods. *International Journal of Computer Science and Information Technologies*, **6**, 3119–3124.
- [4] Witten, I. H. dan Alistair Moffat, T. C. B. (1999) Managing gigabytes. Bagian dari Mann, J. (ed.), *Managing Gigabytes*. Morgan Kaufmann Publishers, London.
- [5] Dash, M. D. R. A. K. . A. M., R. (2010) A hybridized k-means clustering. *International Journal of Engineering, Science and Technology*, **2**, 59–66.
- [6] Tripathi, D. dan Jena, U. (2016) Cognitive and social information based pso. *International Journal of Engineering, Science and Technology*, **8**, 64–75.
- [7] Nadjet Kamel, I. O. dan Baali, K. (2013) A sampling pso-k-means algorithm for document clustering. *Proceedings of 7 th International Conference on Genetic and Evolutionary Computing*, Prague, Czech Republic, 25-27 August, pp. 45–54. Springer, Germany.