

BAB 6

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Dari penelitian yang telah dilakukan, didapatkan beberapa buah kesimpulan yaitu:

1. Reduksi Dimensi pada VSM dilakukan dengan menerapkan *Singular Value Decomposition* yang merupakan teknik dalam metode *Latent Semantic Indexing*. Reduksi Dimensi dilakukan dengan menentukan *rank* atau jumlah dimensi matriks yang tersisa yang didapatkan dari jumlah nilai *singular value*, sehingga akan terbentuk matriks baru yang didapatkan dari perkalian matriks *document-term* dengan *transpose* matriks V yang telah direduksi berdasarkan nilai *rank*.
2. *Text clustering* dengan menggunakan *Fuzzy C-Means* dilihat berdasarkan derajat keanggotaan setiap data pada setiap *cluster*, sedangkan *Text clustering* dengan menggunakan *K-Means* dilihat berdasarkan jarak antara titik data dengan titik pusat *cluster*.
3. Cara untuk membandingkan metode *Fuzzy C-Means* dan *K-Means* adalah dengan membangun perangkat lunak *Text Clustering* yang mengimplementasikan metode LSI, *Fuzzy C-Means*, dan *K-Means*.
4. Secara performa LSI-FCM bekerja lebih baik dibanding LSI-KMeans. Dilihat berdasarkan *accuracy* dari algoritma LSI-FCM lebih besar dibanding *accuracy* algoritma LSI dan *K-Means*. Selain itu, F1 dari algoritma LSI-FCM lebih besar dibanding algoritma LSI dan *K-Means* yang menandakan ketepatan dan keberhasilan informasi yang diminta oleh pengguna dengan jawaban yang dikembalikan program dengan menggunakan algoritma LSI dan FCM lebih besar dibanding dengan menggunakan algoritma LSI dan *K-Means*.
5. Secara kecepatan LSI-FCM bekerja lebih lambat dibanding LSI-KMeans dikarenakan algoritma FCM memiliki serangkaian proses matematis yang lumayan panjang seperti menghitung dan memperbaharui nilai derajat keanggotaan masing-masing data, menghitung pusat *cluster*, dan menghitung fungsi objektif.
6. Hasil *text clustering* dengan menggunakan algoritma LSI-FCM setiap anggotanya memiliki jarak yang lebih dekat dengan titik pusat *cluster*nya dibanding dengan menggunakan algoritma LSI dan *K-Means*.

6.2 Saran

Saran untuk penelitian selanjutnya adalah sebagai berikut.

- Pada penelitian ini, program berhasil dibuat dan bekerja dengan cukup baik. Namun dalam pengujian performa mengalami kendala yaitu proses *text mining* yang cukup lama jika menggunakan dataset dengan jumlah yang cukup besar. Hal tersebut disebabkan oleh pembobotan *term* untuk membentuk matriks *document-term* yang bertipe *array* yang memiliki kompleksitas $O(N^2)$.

- Dataset yang digunakan masih sangat sedikit karena hanya berjumlah 25 *file* .txt. Hal tersebut dikarenakan set data yang terlalu besar rawan menyebabkan perangkat lunak tidak merespon.
- Saran untuk penelitian ini adalah menggunakan struktur data selain *array*. Selain itu, pengujian yang dilakukan dapat menggunakan set data yang lebih besar.

DAFTAR REFERENSI

- [1] Manning, C. D., Raghavan, P., dan Schotze, H. (2009) *An Introduction of Information Retrieval*, 1st edition. Cambridge UP, Cambridge, England.
- [2] Kou, G. dan Peng, Y. (2015) An application of latent semantic analysis for text categorization. *International Journal of Computers Communications & Control*, Chengdu, China, June, pp. 357–369. Springer.
- [3] Hasan, M. M. dan Matsumoto, Y. (2013) Document clustering: Before and after the singular value decomposition. Technical Report UU-CS-2013-007. Nara Institute of Science and Technology, Nara, Japan.
- [4] Hidayati, N., Jambak, M. I., dan Saputra, D. M. (2017) Pengaruh singular value decomposition terhadap metode-metode clustering. Technical Report A15. Universitas Sriwijaya Palembang, Palembang, Indonesia.
- [5] Cline, A. K. dan Dhillon, I. S. (2007) Computation of the singular value decomposition. Bagian dari Hogben, L. (ed.), *Handbook of Linear Algebra*. Chapman & Hall/CRC, Boca Raton.
- [6] Yang, N., Li, S., Sun, R., dan Yang, Y. (2018) Text classification methods based on svd and fcm. *Web and Big Data*, Macau, China, 23-25 July, pp. 111–120. Springer.