

BAB 6

KESIMPULAN DAN SARAN

Pada bab ini akan dibahas kesimpulan seluruh penelitian skripsi dan saran pengembangan untuk penelitian ini

6.1 Kesimpulan Penelitian

Berdasarkan hasil penelitian pada skripsi ini didapatkan beberapa kesimpulan sebagai berikut.

1. Telah diimplementasikan deteksi komunitas menggunakan teknik *Motif Finding*. Lalu, analisis profil komunitas dengan *Co-Occurrence Networks* untuk *directed big graph* yang berasal dari data Twitter. Sebagian besar proses telah memanfaatkan teknologi *big data* Apache Spark dan *library* GraphFrames;
2. Pada penelitian ini diusulkan suatu pendekatan analisis profil komunitas, untuk menjawab pertanyaan seperti “topik apa yang diminati anggota komunitas?”, atau “mengapa komunitas tersebut bisa terbentuk?”. Pendekatan yang diusulkan adalah dengan menganalisis teks *tweet* yang dikirimkan pada komunitas menggunakan metode pemodelan topik. Metode tersebut terbukti dapat mencari kata yang mendeskripsikan profil komunitas terkait. Namun, masih perlu pengembangan lebih lanjut agar hasil yang didapatkan lebih valid dan optimal.

6.2 Saran Pengembangan

Berdasarkan hasil penelitian pada skripsi ini, terdapat beberapa saran untuk pengembangan penelitian selanjutnya, sebagai berikut.

1. Melakukan *text pre-processing* tambahan, seperti lematisasi, *POS tagging*, *named entity recognition* (NER), *parsing*;
2. Menggunakan teknik pemodelan topik lain, seperti *Latent Dirichlet Allocation* (LDA) atau *Non-Negative Matrix Factorization* (NMF);
3. Mencoba metode teknik yang diusulkan untuk domain lain yang sedang tren dan viral, seperti domain investasi atau hukum di Indonesia atau untuk bahasa lain, seperti bahasa Inggris;
4. Mengganti metode analisis profil komunitas dengan opsi yang lebih baik, selain pemodelan topik;
5. Menangani data *spam* saat pengumpulan data, sehingga data teks menjadi lebih bersih;
6. Metode yang diusulkan pada penelitian ini dapat dikembangkan lebih lanjut, seperti analisis *Co-Occurrence* data teks dapat diterapkan pada analisis data Twitter secara *real time*. Contohnya untuk melihat pembicaraan yang sedang tren dan viral di Twitter, lewat kemunculan kata-katanya. Agar hasil analisis lebih menarik, dapat diintegrasikan dengan visualisasi interaktif dan *word cloud*;
7. Selain itu, metode pemodelan topik untuk menganalisis profil komunitas juga dapat diterapkan pada kasus lain, seperti pada komunitas yang terbentuk di forum diskusi *game* atau forum penggemar *gadget*. Komentar-komentar pada forum tersebut berpotensi untuk dianalisis lebih lanjut agar mendapat informasi berharga. Nantinya informasi tersebut dapat dimanfaatkan

untuk rekomendasi, seperti rekomendasi produk atau rekomendasi kepada pengguna dengan minat sama yang belum bergabung ke forum;

8. Mengembangkan arsitektur pengumpul data dengan cara mengotomatisasi proses *extract, transform, load (ETL)*, hingga data mentah (*raw data*) siap dianalisis. Opsi lainnya dengan mengintegrasikan proses *ETL* menggunakan teknologi *cloud* agar lebih *scalable* dan fleksibel.

DAFTAR REFERENSI

- [1] Rosen, K. H. (2002) *Discrete Mathematics and Its Applications*, 7th edition. McGraw-Hill Higher Education, New York.
- [2] Wills, J., Owen, S., Laserson, U., dan Ryza, S. (2015) *Advanced Analytics with Spark: Patterns for Learning from Data at Scale*, 2nd edition. O'Reilly Media, Inc., USA.
- [3] Zafarani, R., Abbasi, M. A., dan Liu, H. (2014) *Social Media Mining: An Introduction*. Cambridge University Press, United Kingdom.
- [4] Karya, G. (2020) Teknologi big data. Bagian dari Moertini, V. S. dan Adithia, M. T. (ed.), *Pengantar Data Science dan Aplikasinya bagi Pemula*. Unpar Press, Bandung.
- [5] Karau, H., Konwinski, A., Wendell, P., dan Zaharia, M. (2015) *Learning Spark: Lightning-Fast Big Data Analytics*, 2nd edition. O'Reilly Media, Inc., USA.
- [6] Chambers, B. dan Zaharia, M. (2018) *Spark: The Definitive Guide Big Data Processing Made Simple*. O'Reilly Media, Inc., USA.
- [7] Buyya, R., N. Calheiros, R., dan Vahid Dastjerdi, A. (2016) *Big Data: Principles and Paradigms*. Morgan Kaufmann Publishers Inc., USA.
- [8] Moertini, V. S. dan Adithia, M. T. (2021) Uncovering active communities from directed graphs on distributed spark frameworks, case study: Twitter data. *Big Data and Cognitive Computing*, **5**.
- [9] Needham, M. dan E. Hodler, A. (2019) *Graph Algorithms*. O'Reilly Media, Inc., USA.
- [10] Erl, T., Khattak, W., dan Buhler, P. (2016) *Big Data Fundamentals: Concepts, Drivers Techniques*. Prentice Hall Press, USA.
- [11] Karau, H. dan Warren, R. (2017) *High Performance Spark: Best Practices for Scaling and Optimizing Apache Spark*. O'Reilly Media, Inc., USA.
- [12] Dave, A., Jindal, A., Li, L. E., Xin, R., Gonzalez, J., dan Zaharia, M. (2016) Graphframes: An integrated api for mixing graph and relational queries. *Proceedings of the Fourth International Workshop on Graph Data Management Experiences and Systems*, New York, NY, USA. Association for Computing Machinery.