

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan penelitian yang telah dilakukan untuk implementasi algoritma K-Means dan GMM untuk automasi pengelompokan berita di kompas.com, dapat ditarik kesimpulan sebagai berikut:

1. Algoritma K-Means dan GMM untuk automasi pengelompokan berita di kompas.com telah berhasil diimplementasikan. Masing-masing dari algoritma *clustering* melakukan pengelompokan dimulai dari 4 *cluster* hingga 10 *cluster*. Hyperparameter yang digunakan untuk algoritma K-Means yaitu dengan `random_state 52`, sedangkan untuk algoritma GMM yaitu dengan `random_state 42`, dan `covariance_type tied`. Implementasi algoritma K-Means dan GMM berhasil dibuat dengan menggunakan *library* Scikit learn yang ditulis menggunakan bahasa pemrograman Python dengan menggunakan Google Colab dan Jupyter Notebook.
2. Berdasarkan hasil evaluasi, dapat disimpulkan bahwa algoritma *clustering* GMM lebih unggul dalam tingkat akurasi dan kecepatan dibandingkan algoritma K-Means dengan tingkat akurasi tertinggi 0.7994 dan kecepatan 2.3722 detik untuk 4 Cluster, 0.7614 dan kecepatan 2.4163 detik untuk 5 Cluster, 0.7428 dan kecepatan 2.9533 detik untuk 6 Cluster, 0.7249 dan kecepatan 2.5499 untuk 7 Cluster, 0.7267 dan kecepatan 3.2291detik untuk 8 Cluster, 0.7114 dan kecepatan 3.3909 detik untuk 9 Cluster, dan 0.6726 dan kecepatan 3.3541 untuk 10 Cluster.

5.2 Saran

Berdasarkan penelitian yang sudah dilakukan, berikut merupakan beberapa saran yang mungkin bermanfaat untuk penelitian berikutnya.

1. Meningkatkan stopwordlist dan teknik *filtering words* untuk mem-filter kata yang tidak diperlukan dan memperkuat identitas suatu berita.
2. Memperbanyak jumlah *dataset* agar algoritma *clustering* dapat melakukan pengelompokkan dengan lebih baik.
3. Mengenali karakteristik *dataset* untuk memilih algoritma *scaling* data dan *dimensionality reduction* yang sesuai
4. Melakukan beberapa perubahan pada hyperparameter yang digunakan, seperti `covariance_type`, `random_state` dan sebagainya.