

Gelombang 1 [55 poin]

Jawab dengan se jelas-jelasnya.

Tidak akan ada tanya-jawab terkait dengan soal. Silakan ambil dan tuliskan asumsi masing-masing (jika diperlukan) dalam menjawab soal.

A. Benar/Salah - jelaskan apakah pernyataan di bawah ini benar/salah, dan berikan alasan anda. [0.5 poin untuk jawaban benar/salah yang tepat, 3 poin untuk alasan yang tepat, total 3.5 poin masing-masingnya, total keseluruhan 17.5 poin]

1. Clustering dapat digunakan untuk **menaikkan ranking** suatu dokumen yang sebetulnya **kurang relevan** secara tekstual dengan query yang diberikan.
2. Jika ada **labeled data**, **evaluasi** clustering bisa dilakukan secara **otomatis**.
3. Jika data teks dalam korpus yang ingin di-cluster **banyak**, maka akan lebih **sulit** mendapatkan hasil clustering yang lebih **akurat**.
4. Dalam prakteknya, **k-means clustering** hanya perlu dijalankan **1 kali hingga konvergen** untuk mendapatkan cluster terbaik.
5. **Minibatch k-means clustering** lebih **cepat** dibandingkan **k-means clustering** biasa.

B. Esai

1. [17.5 poin] Misalkan komputer anda tidak cukup kuat untuk menjalankan algoritma k-means clustering pada suatu korpus tertentu hingga betul-betul konvergen dengan akurat.
 - a. [4.5 poin] Tanpa mengganti/memanipulasi korpusnya dan tanpa meng-upgrade PC anda, sebutkan **tiga** cara agar anda tetap bisa menjalankan algoritma tersebut dan mendapatkan hasilnya! **Note**: hasilnya tidak harus yang paling akurat.
 - b. [9 poin] Dari tiga cara di atas, jelaskan masing-masingnya mengapa hal tersebut bisa membantu algoritma anda tetap berjalan!
 - c. [4] Jika anda diminta untuk memanipulasi/mentransformasi dokumen-dokumen di dalam korpus terlebih dahulu sebelum di-*cluster* sehingga algoritma k-means lebih mudah/cepat dijalankan dan hasilnya bisa diperkirakan masih relatif akurat, apa yang bisa anda dilakukan? **Hint**: pikirkan dokumen seperti apa yang lambat diproses.

