

**Catatan terkait dengan tahapa evaluasi**

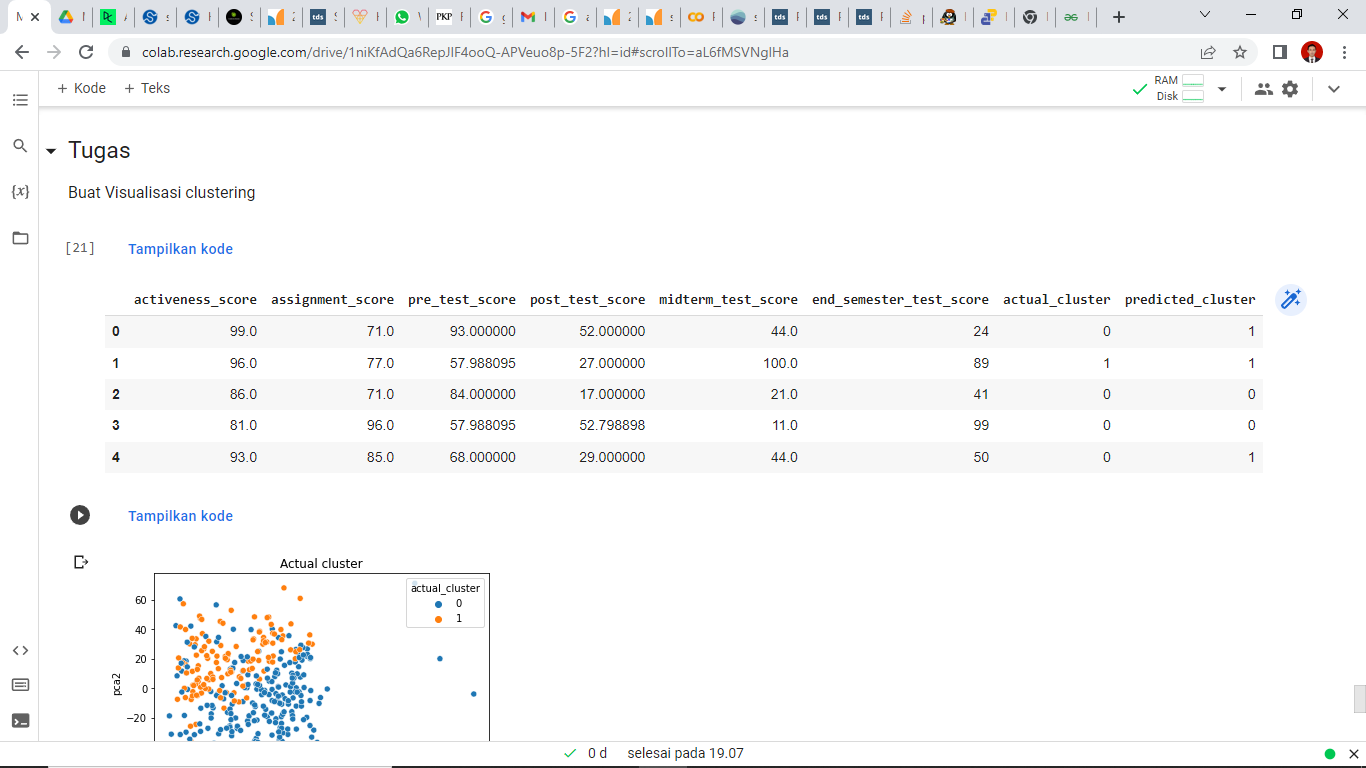
Pemilihan metrik untuk evaluasi model didasarkan pada **data** (karakteristik) dan **tipe permasalahan ML** (clustering/classification/regression)

Contoh:

* Untuk permasalahan classficiation, metrik yang digunakan berdasarkan pada confusion matrix
* Untuk permasalahan regression, metrik yang dapat digunakan antara lain mean squared error (MSE), mean absolute error (MAE), dll
* Untuk clustering, metrik yang dipilih bisa juga tergantung datanya, jika tidak ada cluster yang sudah dilabeli, maka bisa menggunakan metriks silhouette score, dll, jika ada cluster yang sudah dilabeli bisa menggunakan adjusted rand index dll

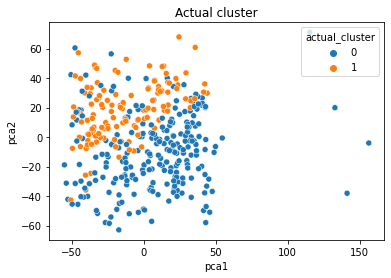
Referensi:

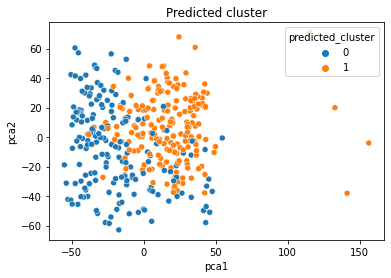
<https://towardsdatascience.com/performance-metrics-in-machine-learning-part-3-clustering-d69550662dc6>



\*actual\_cluster didapat dari kolom pass\_test (not pass = 0, selain itu 1)

\*predicted\_cluster didapat dari labeling hasil pemodelan agglomerative clustering





\*visualisasi style grafik tidak harus sama