



## INTISARI

### ALGORITMA KAMILA UNTUK ANALISIS *CLUSTERING* PADA DATA TIPE CAMPURAN

oleh

Hanny Nur Hanifah

15/383357/PA/17017

Data yang tersedia untuk diolah dapat terdiri dari campuran data bertipe kontinu dan kategorik. Pengelompokkan data tipe campuran tersebut dapat dilakukan dengan beberapa metode. Namun beberapa metode memerlukan pembobotan yang tepat antara variabel kontinu dan variabel kategorik, serta perlunya asumsi parametrik yang kuat. Algoritma KAMILA merupakan gabungan dari metode algoritma *k-means* dan Gaussian-multinomial *mixture models*, dimana nantinya algoritma KAMILA dapat menangani kelemahan dari dua metode tersebut. Selanjutnya dengan menggunakan nilai *Adjusted Rand Index*, akan ditunjukkan bahwa algoritma KAMILA lebih baik dalam mengelompokkan data dibandingkan metode *weighted k-means*.

Kata Kunci: *clustering*, KAMILA, data tipe campuran, *k-means*, Gaussian-multinomial *mixture models*



## ABSTRACT

### *KAMILA ALGORITHM FOR MIXED-DATA CLUSTERING ANALYSIS*

by

Hanny Nur Hanifah

15/383357/PA/17017

Datasets that available to process consist of mixed continuous and categorical data. There are several methods can be used for mixed data clustering analysis. We show that current clustering method need appropriate weighting between continuous and categorical data, also need strong parametric assumptions. KAMILA algorithm combines two popular methods,  $k$ -means algorithm and Gaussian-multinomial mixture models, that can handle their weakness. We show that KAMILA algorithm is better than weighted  $k$ -means method from their Adjusted Rand Index value.

Key Words: clustering, KAMILA, mixed data,  $k$ -means, Gaussian-multinomial mixture models