

## Abstrak

Permasalahan yang muncul dalam suatu data ketika data tersebut terdiri dari atribut numerik dan kategori. Untuk mengatasi masalah tersebut diperlukan teknik khusus untuk melakukan *clustering* pada data yang terdiri oleh atribut numerik dan kategori. Salah algoritma *clustering* yang digunakan untuk melakukan *clustering* pada campuran data numerik dan kategori adalah *Semi-supervised Regression Model*. Proses *clustering* dilakukan dengan menggabungkan *multiple linear regression* untuk data numerik dan *k-mode clustering* untuk data kategori. Data dikelompokkan berdasarkan nilai paling kecil dari *least square error* untuk atribut numerik dan *dissimilarity measures* untuk atribut kategori terhadap pusat suatu *cluster*. Hasil yang diperoleh adalah algoritma *semi-supervised regression model* cocok untuk diaplikasikan pada data yang memiliki atribut numerik dan kategori dimana pada data tersebut rentang nilai pada atribut numerik tidak terlalu jauh serta nilai standar deviasi yang kecil dan pada data kategori tidak memiliki persebaran suatu nilai yang sama untuk *cluster* berbeda.

**Kata kunci :** *clustering, semi-supervised, k-mode, multiple regression, least square error, dissimilarity measure*