

Abstrak

Pada dataset yang sangat besar memungkinkan ada data yang tidak lengkap atau datanya tidak ada sama sekali, disebut dengan *missing values*. Untuk beberapa kasus, *missing values* terkadang muncul tidak secara eksplisit, namun muncul sebagai data yang berpotensi dianggap valid atau muncul sebagai *inliers*. Kejadian seperti inilah yang disebut dengan *disguised missing data*. Adanya *disguised missing data* akan mempengaruhi hasil akhir dari analisis data yang dilakukan, semisalnya pada proses *data mining*.

Pada saat ini, kasus *disguised missing data* sering ditangani dengan menggunakan *outlier detection* atau dengan *anomaly detection*. Namun untuk kasus *disguised missing data* yang muncul sebagai *inliers*, *ourlier detection* dan *anomaly detection* tersebut tidak bekerja dengan baik. Maka dari itu diperlukan suatu pendekatan yang lebih handal. *Embedded unbiased sample* (EUS) *heuristic* merupakan pendekatan yang efektif untuk memecahkan permasalahan ini, karena pendekatan ini dapat menemukan *disguised values* yang frekuen atau yang sering muncul. Dengan menggunakan *Correlation-Based Sample Quality Score* (CBSQS) dan *Chi-Square 2 Sample Test* (CS2ST), EUS *heuristic* akan lebih efisien.

Kata kunci : *Data Mining, Disguised Missing Data, EUS Heuristic, CBSQS, CS2ST*