Abstract

In large datasets there will possibly appears incomplete data or no data at all, it is known as missing values. For some cases, missing values sometimes appears not explicitly, but appears as potentially valid data or appears as inliers. Such occurrence known as disguised missing data. Disguised missing data will impair our result in data analysis, as example at data mining process.

Nowadays, disguised missing data usually handle with outlier detection or with anomaly detection. But for disguised missing data that appears as inliers, outlier detection and anomaly detection can not work well. For that reason, we need a better approach. Embedded unbiased sample (EUS) *heuristic is an effective approach to tackle this problem, because this approach can finds frequent disguised values. With Correlation-Based Sample Quality Score* (CBSQS) *and Chi-Square 2 Sample Test* (CS2ST), EUS *heuristic will be more efficient.*

Keyword : *Data Mining, Disguised Missing Data, EUS Heuristic, CBSQS, CS2ST*