Framework untuk mendeteksi pemalsuan data pada mobile survey

Ibnu Santoso

Abstrak

Interviewer falsifications are relevant problem faced by institutions conducting census and surveys around the world, including BPS-Statistics Indonesia. Falsified data may cause serious impact to generated statistics even though the proportion of falsified data is very small. Usage of Computer Assisted Personal Interviewing (CAPI) in field data collection has proven to improve efficiency and effectiveness. In addition, the use of CAPI is believed to be able to detect data falsification better. This is because CAPI devices can produce a variety of metadata that can not be obtained when using paper questionnaires. This study discusses relevant features to detect interviewer falsification in CAPI-based surveys, validates them, and uses them to identify interviewer falsification automatically using data mining techniques so that human supervisors can take further actions. After analyzing relevant features and conducting experiment, the result showed that unsupervised classification algorithm using simple 2-means clustering could have up to 70,5% accuracy, while supervised classification using logistic regression could have up to 88,5% accuracy.