This is the published version of a paper presented at *2nd International Conference in Accounting Researchers and Educators (ICARE 2016), Sri Lanka, 11th January 2016.*.

N.B. When citing this work, cite the original published paper.

Data Deluge and Its Analysis Issues

Wijayatunga, P.[1]

Current availability of enormous amount of data is mainly due to technological advances. They are useful drawing inferences for creating new businesses, formulation of new policies or revising existing ones, etc. However, much of analyses are performed either by subject domain experts implementing mathematical and computational models incorrectly or by mathematical and computational professionals, purely on data driven basis without paying required attention to the subject domain knowledge. Both of these exercises often result in incorrect inferences and therefore they may harm the society, especially when their inferences are used in practice. We argue that, in order to get valid inferences these two parties should work together. Here we briefly discuss some of the issues that the large-scale data analyses should take into account, especially in open data and big data. We also briefly discuss our solutions that are rather simple to implement.

Keywords: Open data, Big data, Statistical, Causal, Inference

[1]Department of Statistics, Umeå School of Business and Economics, Umeå University Sweden. (priyantha.wijayatunga@umu.se)