Data Fraud Detection

Hans-J. Lenz
Freie Universität Berlin, Germany

Abstract: Data Fraud is a criminal activity done by at least one person who intentionally acts secretly to deprive other people of something of value, for their own benefit, i.e. profit or prestige. Data Fraud happened and still happens everywhere in all centuries and in all fields of activities of human mankind: Business, economics, politics, science, health care, religious communities etc. Data fraud is extensionally characterized by four fields: Data scout, plagiarism, manipulation and fabrication. Data scout is spying out data like NSA and others secret services in GB, Russia etc. globally are doing. Data Plagiarism suppresses referring to the source or provenance of data used by the deceiver. Data Manipulation takes existing data and manipulates (“fine-tuning”) the content encapsulated in tables, diagrams or (historical) pictures where mostly numbers of all data types are manipulated. Finally, Data Fabrication generates artificial data in a brute force way thereby avoiding expensive data recording, time-consuming observations or running statistically well-planned experiments. There is no and will be no omnibus test available to detect data fraud of all kinds. However, a bundle of techniques is available like substring and metadata matching, probability / frequency distribution dependent methods, Benford’s Law, multivariate inliers and outlier tests as well as tests of conformity between a given data set and a fully specified model. The main objective is to give hints to data fraud betrayers with low rates of false positives and negative cases. Some famous historical and actual cases and a bundle of useful tests are presented.

BRIEF BIOGRAPHY

In 1973 Hans-J. Lenz got a doctoral degree (PhD like) in Statistics and Operations Research, Freie Universität Berlin, Germany. In 1978 he was offered a Professorship of Applied Computer Science and Statistics at Freie Universität Berlin, and one of Statistics at University of Bonn. He accepted the first one, and retired there in 2008. His present research concerns Business Intelligence, data quality assessment at the business and economics level in co-operation with the Institute of North America Studies, Berlin, data fraud detection, model-based controlling under uncertainty, and cost/benefit and risk calculations of oil / gas exploration and production in cooperation together with Technical University Berlin.

His research activities in 1969 – 2008 led to ~ 25 books published or co-edited, ~ 350 technical papers authored or co-authored and ~ 25 PhD students supervised.

He received a honorary membership from the Romanian Statistical Society in 2005, and in he got the Golden Medal of Freie Universität Berlin for excellence of service.