



IPS Paper

Data Science in Statistics: methodological and applied issues

Author: Prof. Elisabetta Carfagna

Submission ID: 1038

Reference Number: 1038

Brief Description

Data science has a great and increasing importance in several branches of statistics using large data sets and new data sources, e.g., administrative registers, satellites and aircrafts, webcams, data voluntarily provided by internet users, data harvested from the web and so on.

The analysis and elaboration of these kinds of data require the use of data science methods and tools besides “traditional” statistical methods.

The applications of data science tools range from earth observation to official statistics, and the discussion on advantages, disadvantages, limitations, and requirements of the use of alternative data sources integrated with probability sample surveys is informing the debate in national and international statistical systems all over the world. This Invited Paper Session (IPS) focuses on most relevant methodological and applied issues of data science: interpretability of machine learning tools, potential bias, integration of new data sources with sample surveys for improving official statistics, analysis of huge amounts of meteorological and remote sensing data. This IPS is proposed by the vice-chair and chair-elect of the ISI Special Interest Group on Data Science, discusses methodological and applied issues, is balanced from geographical and gender point of view.

Abstract

Data science has a great and increasing importance in several branches of statistics using large data sets and new data sources, e.g., administrative registers, satellites and aircrafts, webcams, data voluntarily provided by internet users, data harvested from the web and so on.

The analysis and elaboration of these kinds of data require the use of data science methods and tools besides “traditional” statistical methods.

The applications of data science tools range from earth observation to official statistics, and the discussion on advantages, disadvantages, limitations, and requirements of the use of alternative data sources integrated with probability sample surveys is informing the debate in national and international statistical systems all over the world.

This Invited Paper Session (IPS) focuses on most relevant methodological and applied issues of data science: interpretability of machine learning tools, potential bias, integration of new data sources with sample surveys for improving official statistics, analysis of huge amounts of meteorological and remote sensing data.

This IPS is proposed by the vice-chair and chair-elect of the ISI Special Interest Group on Data Science, discusses methodological and applied issues, is balanced from geographical and gender point of view.