



IPS Paper

Data Science Structure and the role of Statistics

Author: Prof. Elisabetta Carfagna

Submission ID: 1404

Reference Number: 1404

Presentation File

abstracts/ottawa-2023_206bdd5ca43a4eb997c5f626bb714297.pdf

Files/Uploads

[IPS 421 Carfagna Elisabetta](#)

Brief Description

Data science has emerged as a very strong, visible, and publicly recognized label for problem-solving using ever-growing, large datasets and new data sources such as administrative registers, satellites and aircrafts, webcams, data voluntarily provided by the internet users, data harvested from the web and so on.

The applications of data science tools range from earth observation to official statistics.

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.

The crucial question is: what is "data science"? Is it statistics or computer science? Which is its relationship with "big data", "new data sources", "machine learning", "artificial intelligence" and "smart statistics"? These questions will be addressed in this paper.

Abstract

Data science has emerged as a very strong, visible, and publicly recognized label for problem-solving using ever-growing, large datasets and new data sources such as administrative registers, satellites and aircrafts, webcams, data voluntarily provided by the internet users, data harvested from the web and so on. The applications of data science tools range from earth observation to official statistics. 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. The crucial question is: what is "data science"? Is it statistics or computer science? Which is its relationship with "big data", "new data sources", "machine learning", "artificial intelligence" and "smart statistics"? These questions will be addressed in this paper.