Chapter 8

Data Science: Recent Developments and Future Insights:

Sabitha Rajagopal
SNS College of Technology, Anna University, India

ABSTRACT

Data Science employs techniques and theories to create data products. Data product is merely a data application that acquires its value from the data itself, and creates more data as a result; it’s not just an application with data. Data science involves the methodical study of digital data employing techniques of observation, development, analysis, testing and validation. It tackles the real time challenges by adopting a holistic approach. It ‘creates’ knowledge about large and dynamic bases, ‘develops’ methods to manage data and ‘optimizes’ processes to improve its performance. The goal includes vital investigation and innovation in conjunction with functional exploration intended to notify decision-making for individuals, businesses, and governments. This paper discusses the emergence of Data Science and its subsequent developments in the fields of Data Mining and Data Warehousing. The research focuses on need, challenges, impact, ethics and progress of Data Science. Finally the insights of the subsequent phases in research and development of Data Science is provided.

1. INTRODUCTION

Data science employs techniques and theories drawn from many fields within the broad areas of mathematics, statistics, operations research, information science, and computer science, including signal processing, probability models, machine learning, statistical learning, data mining, database, data engineering, pattern

DOI: 10.4018/978-1-5225-3015-2.ch008
Data Science

recognition and learning, visualization, predictive analytics, uncertainty modeling, data warehousing, data compression, computer programming, artificial intelligence, and high performance computing. Methods that scale to big data are of particular interest in data science, although the discipline is not generally considered to be restricted to such big data, and big data solutions are often focused on organizing and preprocessing the data instead of analysis. The development of machine learning has enhanced the growth and importance of data science (Davenport and Patil, 2012; “Wikipedia Data Science” n.d.) Data science affects academic and applied research in many domains, including machine translation, speech recognition, robotics, search engines, digital economy, but also the biological sciences, medical informatics, health care, social sciences and the humanities. It heavily influences economics, business and finance. From the business perspective, data science is an integral part of competitive intelligence, a newly emerging field that encompasses a number of activities, such as data mining and data analysis (Hays, 2004). This paper discusses the emergence of Data Science and its subsequent developments in the fields of Data Mining and Data Warehousing. The research focuses on need, challenges, impact, ethics and progress of Data Science. Finally the insights of the subsequent phases in research and development of Data Science is provided.

2. DATA SCIENCE AND ITS EMERGENCE

The term “data science” has existed for over thirty years and was used initially as a substitute for computer science by Peter Naur in 1960. In 1974, Naur published a Concise Survey of Computer Methods, which freely used the term data science in its survey of the contemporary data processing methods that are used in a wide range of applications. In 1996, members of the International Federation of Classification Societies (IFCS) met in Kobe for their biennial conference. Here, for the first time, the term data science is included in the title of the conference (“Data Science, classification, and related methods”).

In November 1997, C.F. Jeff Wu gave the inaugural lecture entitled “Statistics = Data Science?” for his appointment to the H. C. Carver Professorship at the University of Michigan. In this lecture, he characterized the statistical work as a trilogy of data collection, data modeling and analysis, and decision making. In his conclusion, he initiated the modern, non-computer science, usage of the term “data science” and advocated that statistics be renamed data science and statistician data scientists. Later, he presented his lecture entitled “Statistics = Data Science?” as the first of his 1998 P.C. Mahalanobis Memorial Lectures. These lectures honor Prasanta Chandra Mahalanobis, an Indian scientist and statistician and founder of the Indian Statistical Institute.
Virtual Organisational Trust Requirements: Can semiotics help fill the trust gap?
www.igi-global.com/chapter/virtual-organisational-trust-requirements/54438?camid=4v1a

La Terra dei Fuochi: Cultural Labeling, Ecological Crimes, and Social (re)Action in Mediocratic Italy
www.igi-global.com/article/la-terra-dei-fuochi/202473?camid=4v1a