Chapter 1
Introduction to Data Science

M. Govindarajan
Annamalai University, India

ABSTRACT

This chapter focuses on introduction to the field of data science. Data science is the area of study which involves extracting insights from vast amounts of data by the use of various scientific methods, algorithms, and processes. The term data science has emerged because of the evolution of mathematical statistics, data analysis, and big data. Data science helps to discover hidden patterns from the raw data. It enables to translate a business problem into a research project and then translate it back into a practical solution. The purpose of this chapter is to provide emphasis on integration and synthesis of concepts, techniques, applications, and tools to deal with various facets of data science practice, including data collection and integration, exploratory data analysis, predictive modeling, descriptive modeling, data product creation, evaluation, and effective communication.

INTRODUCTION

Data Science is the science and art of using computational methods to identify and discover influential patterns in data. The goal of data science is to gain insight from data and often to affect decisions to make them more reliable (D.Abbott, 2014).

Data is necessarily a measure of historic information so, by definition, data science examines historic data. Considering this definition data science can be defined as organizing data knowledge that can be used for experiments and prediction. The need for data science has developed due to the immense increase in the amount of raw data such as images, text, video and others.

Every field is contributing to this ever increasing data such as engineering, mining, healthcare, hospitality, energy etc. Data scientists are developing various algorithms and techniques in order to process and analyze this data and make the best use out of it. Previously health industry used hard paper to store data regarding patients and other medical issues. But the new trend has helped doctors to store this data in electronic form (Raghupathi & Raghupathi, 2014).

Space exploration has produced a large amount of data considering the recent space missions. Data scientists have also helped to store this data and use it for prediction to carry out further missions. Implementing business logics and strategies need data analysis and predictions which can be now easily done by the use of prediction algorithms.

Energy companies are using data prediction algorithms to manage the energy production based on demands and supply. This prediction has aided to use the available energy efficiently. In coming years machines will have the ability to predict and generate the required resource as per the supply and demand. The current upcoming technology artificial intelligence is also being helped tremendously by the use of data mining and prediction algorithms.

LITERATURE REVIEW

Table 1. Shows various Methods, Tools and Applications used in Data Science

Author	Advantages	Disadvantages
Katerina Lepenioti et al., (2020)	This paper provides clarity on the research field of prescriptiveanalytics, synthesizes the literature review in order to identify the existing research challenges, and outlinesdirections for future research.	The review presented herein is limited to works explicitly scoped as prescriptive analytics. It does not deal with works from different research fields that can potentially contribute to the field of prescriptive analytics.
Ben Kei Daniel (2019)	This paper was inspired partly by insights drawn fromthe literature but mostly informed by experience researching into Big Data in education.	This paper does not discuss privacy, ethics, access and governance issues and identify strategies to support educationalresearchers.
ImanRaeesi Vanani et al., (2019)	This paper discussed the methods of both ML and DL and an ML/DL deployment model for IOT data.	This paper does not discuss the methods of both ML and DL and an ML/DL deployment model for IOT fog and edge computing.
Jeffrey Ray et al., (2018)	This paper provides an overview of the current research efforts in Big Data science, with particular emphasis on its applications, as well as theoretical foundation.	This paper does not consider applications of Natural Language Processing and Network Theory.
Jens Baum et al., (2018)	This paper is based on a broad, systematic literature review consisting of a two-step search approach combined with additional filtering and classification.	This paper considers search strings consisted of the words "big data", "maintenance", "production" and "manufacturing" and did not consider other related terms, such as "factory", "industry 4.0" and so on. This study considered only papers written in English (including the title, abstract and full text); others were excluded.
G. Magesh et al., (2017)	This paper is facilitating future researchers to develop new approaches and algorithms to solve few challenges in big data.	This paper addressed only few challenges in big data and its applications in various fields.
LongbingCao (2017)	This paper provides a comprehensive survey and tutorial of the fundamental aspects of data science.	This paper does not provide algorithms and process of data science.
Yulan Liang et al., (2016)	This paper provides a general survey of recent progress and advances in Big Data science, healthcare, and biomedical research.	This paper does not consider Big Datasets in healthcare, and biomedical research.
SanketMantri (2016)	This paper discussed about data science, current and future development in this field.	This paper does not consider new methods to obtain appropriate values and develop good computational data sets.
PanagiotisBarlas et al., (2015)	This paper provides an overview of open source (OS) data science tools, proposing a classification scheme that can be used to study OS data science software.	The proposed classification scheme provides useful, but limited scope, statistical functionality to the parent tool.

Related Content

Measuring the Quality of Healthcare Services in Bangladesh

Fahima Khanamand Nayem Rahman (2019). *International Journal of Big Data and Analytics in Healthcare (pp. 15-31).*

www.irma-international.org/article/measuring-the-quality-of-healthcare-services-in-bangladesh/232323

Big Data Applications in Vaccinology

Joseph E. Kasten (2021). *International Journal of Big Data and Analytics in Healthcare (pp. 59-80).*

www.irma-international.org/article/big-data-applications-in-vaccinology/276927

Information Measures

(2015). *Formalized Probability Theory and Applications Using Theorem Proving (pp. 129-142).*

www.irma-international.org/chapter/information-measures/127261

Advances in Fuzzy Dynamic Programming

Felix Mora-Camino, Elena Capitanul Conea, Fabio Krykhtine, Walid Moudaniand Carlos Alberto Nunes Cosenza (2018). *Optimization Techniques for Problem Solving in Uncertainty (pp. 116-139).*

www.irma-international.org/chapter/advances-in-fuzzy-dynamic-programming/206632

A Markov-Chain-Based Model for Group Message Distribution in Connected Networks

Peter Bajorskiand Michael Kurdziel (2020). *International Journal of Data Analytics (pp. 13-29).*

www.irma-international.org/article/a-markov-chain-based-model-for-group-message-distribution-in-connected-networks/258918