

The novelty of Data Science and Big Data on society

Prof.(Dr.) Arvind Kumar Upadhyay, CSE-ASET

Amity University Madhya Pradesh Gwalior

E-mail: akupadhyay@gwa.amity.edu

Abstract:

The area of data science deals with collection, preparation and analysis of a huge amount of data that is generated these days. It is not limited to analysis alone but concerns much more and different skills are needed to work in this field. Data architecture, acquisition, analysis, archiving are some of important areas in data science. In the present paper author explores the novelty behind the Data science and its impact on society.

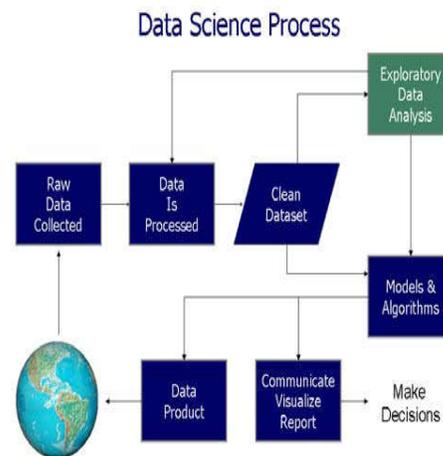
Keywords Data Science, Impact of data science, architecture of data science

1. Introduction

As an important discipline, the Data Science deals with informatics, mathematics and computer sciences. Data science is dependent on statistics and its components. The importance of this field has become even more rigorous after the era of Big Data. Data science can also be linked to Data mining. The fundamentals of Data Mining like query optimisation, fuzzy sets, Bayesian statistics and the like are part of Data Science[1]. So, Knowledge Discovery in Databases is the base on which the concepts of Data sciences are built. The ideas of Data Mining are combined in the notion of Data Science, leading to different definitions. In fact according to one school of thought, researchers utilize the capacity to discover and translate rich information sources; oversee a lot of information

notwithstanding equipment, programming, and transfer speed imperatives; consolidate information sources; guarantee consistency of datasets; make representations to help in comprehension of information; construct scientific models utilizing the information; and display and impart the information experiences/discoveries.

2. Basic steps of data science:



Data science is comprehensive and the fundamentals of this field are: Arranging the data, bundling the data and finally conveying the information[2]. The concept of bundling is inclusive of collection and sorting of data. The field of data science is also concerned with what would be the data, how the data set is prepared, who would use the data and why it is important to the stake holder. The

researcher of this field has to understand about the final result of data science transform and should be clear about the outcome. The limitations of accessible assets and time are detrimental in data science. The stake holder of data science must be well understood. The steps of data science are collecting the data, applying the pre-processing steps, applying the technique to find the patterns in data and finally the knowledge[3].

The Data Science steps

a) Data wrangling and munging

Data wrangling means Data collection from heterogeneous sources and then pre-processing the data so that Data science techniques may be applied for useful purposes[4]. There after next important task is sorting of data and putting prevalent practices in information administration including adjusting the individuals and the frameworks. In bundling, consistency and joining the crude information is maintained.

b) Data Analysis

Data analysis deals with applying Machine learning techniques to pre-processed data sets to extract use information and patterns which contributes to decision making.

c) Convey Data.

This step includes apprising the stake holders of the Data. The conclusion drawn from data analysis is shared with all concerns so that important business decisions may be taken. It empowers the stakeholders of the enterprise to reach next level of awareness and take appropriate actions.

Conclusion And Future scope

This area of Data science emerged out of necessity and it is quite different from research field. It is, in recent years, evolved from being data representation to being useful in engineering and industry. Some prominent areas are as below:

1. Business Analytics –

The exclusive data about business and applying a data science technique may help in making important decisions about business.

2. Prediction –

Predictive models build in data science help in analysis of the data and finding the patterns. Machine learning techniques are useful tools which help realise this concepts.

3. Security –

The output of data analysis may detect fraud and malicious cases of transactions. Financial institutions depend on data mining techniques to detect and prevent frauds.

4. Computer Vision –

Human computer interaction, autonomous vehicles and robotics are some of area in which Computer vision is used. The data is collected from images and video analysis. It enables the computer see and envision.

5. Natural Language Processing –

Machine translation, parsing, sentiment analysis and natural language generation are some of the prominent tasks in Natural language processing. The data in NLP is

collected from corporate sector for analysis.

The field of data science is going to be very demanding; it may find its applications in all kinds of problems. The skill set needed in Data science is varied like visualization, engineering algorithm etc. If the demands of data science are increasing then corporate world would accept diversification etc. The concept of customization may easily be achieved through data science in the fields of healthcare, banking, public services, insurance etc.

Conclusion And Future Work:

As the industry grows day by day, Data Mining and its allied area are becoming important disciplines. As the nature of the data objects are getting more and more complex and arbitrary, this field would continue to grow. The techniques of Data mining are directed towards understanding the behavioural pattern and the current trends in market etc.

Though the area is yet to mature it finds its applications in health care, aerospace, manufacturing etc. It actually helps those areas function well and be more productive.

References:

1. R Agrawal ,T 1 mielinski, A Swami. Database Mining: A Performance Perspective[J]. IEEE Transactions on Knowledge and Data Engineering, 1993,12:914-925.
2. Y. Peng, G. Kou, Y. Shi, Z. Chen (2008). "A Descriptive Framework for the Field of Data Mining and Knowledge Discovery" International Journal of

Information Technology and Decision Making, Volume 7, Issue 4 7: 639 – 682.

doi:10.1142/S0219622008003204.

3. Ming-Syan Chen, Jiawei Han, Philip S yu. Data Mining: An Overview from a Database Perspective[J]. IEEE Transactions on Knowledge and Data Engineering, 1996, 8(6):866-883.
4. Han, J. & M. Kamber, Data mining: concepts and techniques, San Francisco: Morgan Kaufman (2001).