

A REVIEW PAPER ON FORMATION OF DATA SCIENCE AND FUNDAMENTALS

Dnyaneshwar V. Kudande

Department of Computer Engineering
Suman Ramesh Tulsiani Technical Campus-FoE,
Pune, Maharashtra, India.

Bhavesh B. Shah

Department of Computer Engineering
Suman Ramesh Tulsiani Technical Campus-FoE
Pune, Maharashtra, India

Abstract— In this paper, various fields that are related with the formation of data science have been discussed. Data science used techniques like artificial intelligence, machine learning and deep learning to extract useful information and to predict the pattern of future. Data science is used in various application fields like healthcare system, speech recognition, advanced image recognition, search engines, banking sectors and forecasting etc.

Keywords— Artificial Intelligence, machine learning, deep learning

I. INTRODUCTION

Data Science is a scientific disciplinary that influence scientific methods, algorithms, system and processes to extract useful information from many structured and unstructured data. Basically data science introduced using related field like big data and data mining.

In the year of 1996, first time the term ‘Data Science’ were introduced in the title of a statistical conference “International Federation of Classification Societies” (IFCS) known for “Data Science, classification and related methods” [3]. Also the term ‘Data Science’ was determined by the statisticians, in the public image of Data Science and provide the importance of big data analysis.

Next source of Data Science is Knowledge Discovery in Databases (KDD) [5] and its sub topic called Data Mining. In this paper the KDD by now brings collectively forms many approaches for knowledge discovery which includes information theory, expert systems, inductive learning, statistics, fuzzy sets and query optimization. Hence the KDD is a large building block for development and interaction between many different fields for overall objective for finding the knowledge in data.

In the next source of Data Science, these ideas are combined together in the form of notion of Data Science leading towards the different definitions. One of the definitions for Data Science was provided by the Cao in the form of formula [4]:

Data Science = (Informatics + Statistics + Communication + Computing + Management + Sociology) | (Environment + Data + Thinking).

In today’s world, all above ideas are combined together to form the notation of Data Science, using different field. Data Science included with new formula,
Data Science = (Statistics + Artificial Intelligence + Machine Learning + Deep Learning) | (Data + Knowledge + Environment).

II. FORMATION OF DATA SCIENCE

In this section let us discuss about the fundamental formation of data science. Following figure-1 illustrate the structure of all fundamental fields related with the data science [10]

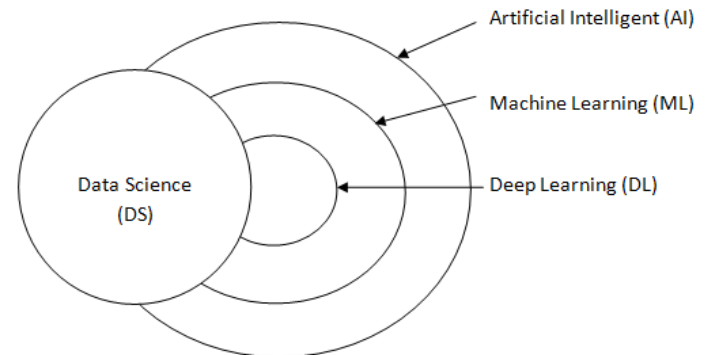


Figure-1: Fundamental fields and formation of data science.

Data science basically consists of following fields:

- i. Artificial Intelligence (AI)
- ii. Machine Learning (ML)
- iii. Deep Learning (DL)

i. Artificial Intelligent (AI):

Artificial Intelligence is the combination of science and engineering for making intelligent machine. [4] In sense of these AI is a technique for making machines to think and behave similar like human being. For example car automation system and android app like Google map are the examples of AI.



ii. Machine Learning (ML):

Machine Learning (ML) is a subset of Artificial Intelligence (AI) which uses statistical tools to explore and analyze the data. Machine Learning (ML) is mainly divided into three types [6]:

Supervised Learning: In this type it uses labeled data and algorithms like Regression and classification etc.

Unsupervised Learning: This type uses unlabeled (data defined without categories or group) data and algorithms like clustering KNN, K-Means etc.

Semi Supervised Learning: The combination of some labeled and some unlabeled data which is called as Semi Supervised Learning. Consider the example Speech Analysis, Internet Content Classification etc. Algorithms used like generative models and graph based models.

iii. Deep Learning (DL):

Deep Learning (DL) is a subset of Machine Learning (ML) which creates architecture like Multi Neural Network to mimic like a human brain. Deep Learning (DL) is mainly divided into three types [6]:

ANN (Artificial Neural Network): Dataset in the form of numbers.

CNN (Convolutional Neural Network): Dataset in the form of images.

RNN (Recurrent Neural Network): Dataset in the form of time series data.

Hence Data Science apply all the technique of Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) using methods like Statistics, Linear algebra, Probability, Differential calculus etc.

III. ANALYSIS AND RESULT

Data Science is a stream of filed that deal with the information analysis from massive amounts of data present in an organization’s warehouse. It includes forming meaningful insights using unstructured and raw data that can be processed using programming, business and analytical skills. In this, we discuss about the data science using python topic in detail. Before started with the topic first we see the scope for python in the field of data science and machine learning software poll in the year from 2017 to 2019 [7].

The following figure-2 shows the different tools of Data Science used by the software industry from the year of 2017 to 2019.

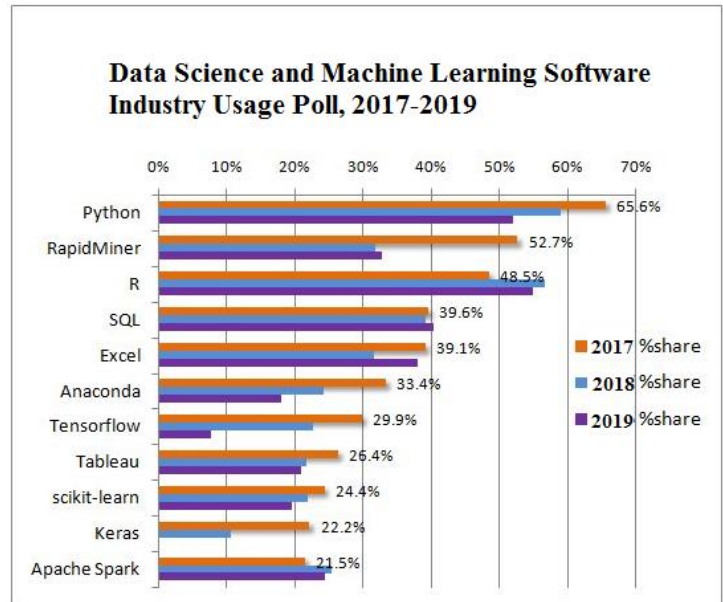


Figure-2 Different Data Science tool usages in Software Industry from year 2017 to 2019

IV. CONCLUSION

This paper covers the fundamental concept and introduction of Data Science. In today’s world each and every person is using data science knowingly or unknowingly. Also this paper focused on the formation of data science. We also believe that data science becomes new type of science, like social science and natural science. In future perspective the formation of data science may updated with the new technologies.

V. ACKNOWLEDGEMENT

This work is supported in part by Suman Ramesh Tulsiani Technical Campus-FoE, Kamshet, Pune which is affiliated to Savitribai Phule Pune University.

VI. REFERENCE

- [1] Cleveland, W. S. (2001) Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics. *International Statistical Review* 69(1), (pp 21–26).
- [2] Cao, L. B. & Yu, P. S. (2009) Behavior Informatics: An Informatics Perspective for Behavior Studies. *IEEE Intelligent Informatics Bulletin* 10(1), (pp 6–11).
- [3] Hey, T., Tansley, S., & Tolle, K. (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery. *Microsoft Research*



- [4] Cao, L. (2017) : Data science: a comprehensive overview, *ACM Computing Survey*.
- [5] Oded Maimon, Lior Rokach (2018): Introduction to knowledge discovery in databases, *Data Mining and Knowledge Discovery Handbook*.
<http://www.ise.bgu.ac.il/faculty/liorr/hbchap1.pdf>
- [6] Castrounis, A. (2017). Artificial Intelligence, Deep Learning, and Neural Networks, Explained. Kdnuggets.com. Available at:
<http://www.kdnuggets.com/2016/10/artificial-intelligence-deep-learning-neuralnetworks-explained.html>
- [7] <https://www.kdnuggets.com/2019/05/poll-top-data-science-machine-learning-platforms.html>
- [8] The Official NVIDIA Blog. (2017). The Difference Between AI, Machine Learning, and Deep Learning? NVIDIA Blog. [online] Available at:
<https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligencemachine-learning-deep-learning-ai/>
- [9] Claus Weihs, Katja Ickstadt (2018), Data Science: the impact of statistics, Springer- *International Journal of Data Science and Analytics*.
- [10] Fionn Murtagh and Keith Devlin (2018), Article The Development of Data Science: Implications for Education, Employment, Research, and the Data Revolution for Sustainable Development, Big Data Cognitive Computing; doi: 10.3390/bdcc2020014.
- [11] Smith, F. Jack (2006) Data Science as an academic discipline. *Data Science Journal* 5, (pp 163–164).
- [12] Zhu, Y. Y. & Xiong, Y. (2011) Dataology and Data Science: Up to Now. Retrieved from the World Wide Web November 16, 2014:
http://www.paper.edu.cn/en_releasepaper/content/4432156
- [13] Iwata, S. C. (2008) Editor’s Note: Scientific ‘Agenda’ of Data Science. *Data Science Journal* 7, (pp 54–56).