

Review Paper on Deep Dissertation of Data Science: Related Issues and Its Applications

Nishant Dandekar¹, Kushal Suvarna², Meena Perla³, Reshma Chaudhari⁴

¹Department of EXTC, Mumbai University, Virar East
Email: nishdan99@gmail.com

²Department of EXTC, Mumbai University, Virar East
Email: kushalsuvarna@viva-technology.org

³Department of EXTC, Mumbai University, Virar East
Email: meenavallakati@viva-technology.org

⁴Department of computer science, Mumbai University, Virar
Email: reshmachaudhari@viva-technology.org

Abstract— In the era of digital evolution data science plays important roll to manage various kind of big data. Data science related to collecting, gathering, performing various operation on data. To handle big data is next to impossible for normal person, but human having ability to generate algorithm to handle such kind of big data. In the given review paper author mainly focus on future scopes and present stage of data science.

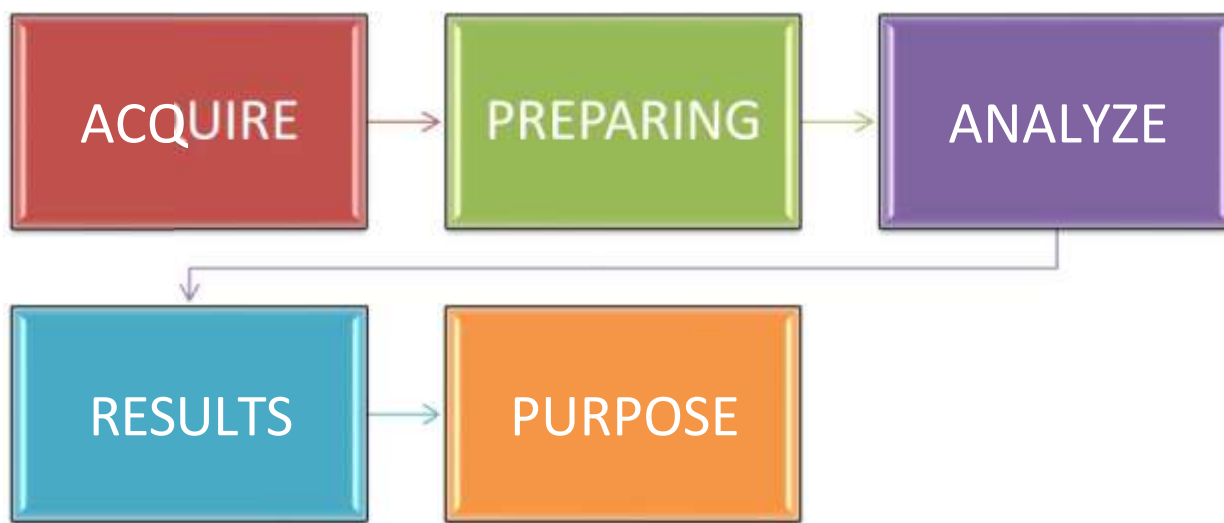
Keywords— Application of Data Science, Data mining, Data science.

1. INTRODUCTION

Now a day's data science used everywhere, the popular field like machine vision, IOT, cloud computing is incomplete without data science. It is thing which used to handle big data which having lots of information. The data science consists of arranging the data, extracting important data from available data, study on available data to store it in less space. Data mining is the important part of data science. It studies previous data and extract useful information out of it. Data science is field which uses scientific method, statistical data, algorithm create fruitful information from past available data or data generated using various sensors.

2. STEPS INVOLVED IN DATA SCIENCE PROCESS

Data science is a vast process which include following steps as shown in block dia.1.



BLOCK DIA.1

2.1 ACQUIRE

Acquire is anything from which we get data including finding accessing or moving data. In the field like IOT we get data from sensors which can be of temperature humidity persons activity, distance etc sensors get all data and stores it in particular memory locations for further use. Acquiring can be finding also we can use data from available source from previous finding. It includes transportation of data from source to distributed file system.

2.2 PREPARE DATA

Prepare data step further classifies into two steps which include explore data and pre-process the data.

Explore data: - As the name suggest it is process in which data scientist or person who using data is just see the available data and try to understand what things we have in the above data, quality and format of the data.

Pre-process data: - The step contain performing basic operation on available data to clean the data to integrate it and make it available in user understandable format. It includes modelling data into more convenient data model or packaging it using specific data format. If more steps are involved the step integrate the data.

2.3 ANALYSIS OF DATA

Prepared data is then pass to analyse data step which include selecting appropriate analysis technique or data model to get desire outcomes. The analytical technique includes hypothesis testing, linear regression, hypothetical regression, neural networks, decision trees etc. this step is completely related to previous step in which format the data is available

2.4 RESULTS

The fourth step contain results which means presenting filter data which is used for different application. The result is in the tabular or statistical format. The step contains summarization or visualization of filter data.

2.5 PURPOSE

The last step brings us back to the first step for which we do data science THE PURPOSE.

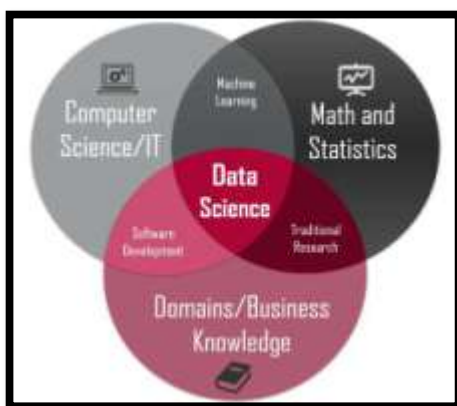


Fig 1

3. APPLICATION

There are innumerable applications of data science which include fraud and risk detection, healthcare, internet search, target advertising, website recommendation. The best application is target application. If you thought Search would have been the biggest of all data science applications, here is a challenger – the entire digital marketing spectrum. Starting from the display banners on various websites to the digital billboards at the railway station – almost all of them are decided by using data science algorithms. This is the reason why digital ads have been able to get a lot higher CTR (Call-Through Rate) than traditional advertisements. They can be targeted based on a user's past behavior.

Augmented reality is the upcoming technology in the field of data science. The popular game Pokemon Go is the best example of both the field. The game contains lots of data of user's location, way of playing the games and according to the articles are placed using augmented reality.

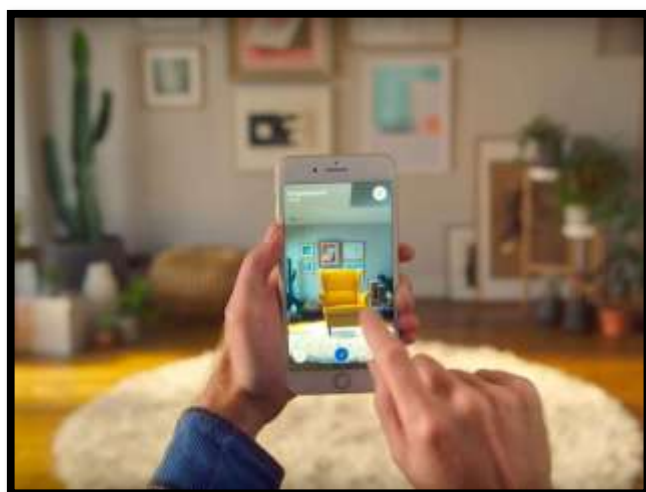


Fig 2: Application Of Data Science (Augmented Reality). Fig 3: Application In Target Advertising

4. CONCLUSION

Now a day's large amount of data is generated through various application. To identify, to process on data and took out useful information is necessary. Data science is a branch of science which related to big data and to perform various operation on it. In this paper, author try to glance on various topic related to data science and its real-life application. Method to create useful and effective data is also explained in this paper. We believe that in future analysts will give careful consideration to these methods to take care of issues of enormous information successfully and effectively.

5. REFERENCES

- [1] "Big Data: The Next Frontier for Innovation Competition and Productivity" in McKinsey Global Institute, 2011.
- [2] B. F. Jones, S. Wuchty, B. Uzzi (2008), in "Multi-University Research Teams: Shifting Impact Geography and Stratification in Science", Science 322, pp. 1259-1262.
- [3] C. L. Philip, Q. Chen and C. Y. Zhang (2014), in Data-intensive applications, challenges, techniques and technologies: A survey on big data, Information Sciences, 275, pp.314-347.
- [4] https://en.wikipedia.org/wiki/Data_science.

- [5] J. Bollen, H. Van, de Sompel, A. Hagberg, R. Chute, M. A. Rodriguez, L. Balakireva (2009), "Clickstream Data Yields High-Resolution Maps of Science", PLoS ONE 4, pp. 1-11.
- [6] J. Dean, S. Ghemawat, Jan (2008), in "MapReduce: Simplified data processing on large clusters", Commun. ACM, vol. 51, no. 1, pp. 107-113.
- [7] J. Manyika, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Byers (2011), from Big data: The Next Frontier for Innovation Competition and Productivity.
- [8] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia, Apr (2010), "A view of cloud computing", Commun. ACM, vol. 53, no. 4, pp. 50-58.
- [9] M. Hilbert, P. Lopez (2011), from "The world's technological capacity to store communicate and compute information", Science, vol. 332, no. 6025, pp. 60-65.
- [10] M. K.Kakhani, S. Kakhani and S. R.Biradar (2015), Research issues in big data analytics, International Journal of Application or Innovation in Engineering & Management, 2(8), pp.228-232. [11] M. M. Waldrop (1992), "Complexity: The Emerging Science at the Edge of Order and Chaos", Simon & Schuster.
- [12] P. Chapman, J. Clinton, R. Kerber, C. Shearer, R. Wirth (2000), in "CRISP-DM 1.0: Step-by-step data mining guide", The CRISP-DM Consortium.
- [13] S. Wuchty, B. F. Jones, B. Uzzi (2007), "The Increasing Dominance of Teams in Production of Knowledge", Science 316, pp. 1038- 1039.
- [14] T. H. Davenport, J. G. Harris (2007), Competing on Analytics: The New Science of Winning, Harvard Business School Press.
- [15] W. v.d. Aalst (2011), Process Mining: Discovery Conformance and Enhancement of Business Processes, Berlin, Germany: Springer-Verlag.