International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VI, June 2018- Available at www.ijraset.com

Introduction to Data Science and Machine Learning

P. Harika Reddy[1] K. Rushikesh[2] Surapaneni Gopi Siva Sai Teja[3]
[1] *Student, Sreenidhi Institute Of Science and Technology, Hyderabad, India*

Abstract: Data Science is the most emerging field of computer science. This paper gives a vast information about how actual data science has been evolving and how machine learning has been changing since ages . And the uses of data science and some brief information about how data science is so powerful and some use cases involving the implementation of data science in the real life.

I. INTRODUCTION

Data Science is interdisciplinary field of scientific methods ,processes ,algorithms and systems to extract knowledge or insights from data in various forms, either structured or unstructured similar to data mining. The term "data science" has appeared in various contexts over the past thirty years but did not become an established term until recently.In an early usage it was used as a substitute for computer science by Peter Naur in 1960. Naur later introduced term "data logy". In 1974, Naur published concise survey of computer methods, which freely used term data science in its survey of contemporary data processing methods that are used in wide range of Applications. In 1996, members of International Federation of classification societies(IFCS) met in Kobe for their biennial conference. Here, for first time the term data science is included in title of conference, after the term introduced in roundtable discussion by Chikio Hayashi Machine learning it's not now we started teaching machines many decades ago we started teaching machines know if as a basic as 2+2 is 4 calculators and computer came into being eventually you know they became starter we got our personalised computer everybody had a computer on their homes then later on IBM developed a computer called deep blue who started beating the grandmasters at chess so computer science started evolving it is still a young discipline but it became our field of study and became a solid evolving then came this term called data science and very soon it became the coolest thing on this earth what is data science data science is basically a multidisciplinary field which brings together the traditional age-old field of statistics and combines it with the recent field of computer science and why is it called as science is because it is more experimental it is empirical it is driven by testing out different algorithms into on real-world problems business problems and it is it starts from getting the data processing it making it ready for use building models on it and those models have always existed in statistics but now those models are starting to drive business outcomes they're starting to drive meaning for businesses to take decisions on and therefore this field has started emerging now data science machine learning they start solving
business problems or problems which we do not even think need to be solved

II. LITERATURE REVIEW

[13] Cathy O'Neil & Rachel Schutt summarize Introduction to Data Science. Also introduces Data Science process ,Statistical Inferences and Algorithms[14] John D Kelleher ,Brendan Tierny summarize what is data science,where is data science used and what is data set and various algorithms for data integration and preparation[15].Lillian Pierson gives abstract of how data science is quickly enlarging field of offering ways for viewing the world around us.And discusses how to work with data science strategies and shares tips to begin career in this field[16]Rachel Schutt and Cathy O'Neil describes doing data science.And found comprehensive approach to data science.[17]

III. WHY DATA SCIENCE?

1.BUSINESS ADMINISTRATION
2. EXPLORATORY DATA ANALYSIS
3. MACHINE LEARNING AND ADVANCED ALGORITHMS
4. DATA PRODUCT ENGINEERING

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VI, June 2018- Available at www.ijraset.com

As you can see from the above table a data analyst usually explains what is going in the processing history of data science, in the other hand data scientist only explains the decisions and predictions making use of predictive casual analytics, perspective analysis

IV. IS DATA SCIENCE SO POWERFUL

Most information in world being shared digitally, and more still being stored on cloud, almost every industry is sitting on treasure trove of data.
If mined properly, such data can reveal information on customer preferences, tastes, usage in real time and much more.
Most organizations still don't know how to use 80% of their data, since most data exists in unstructured format. The key concern is how organizations can use their imagination to pick what data to correlate and analyze.

A. Life Cycle Of A Data Scientist

A data scientist's daily tasks involves around data, which is no surprise given the job title. Data scientists spend much of their time gathering data, looking at data. Shaping data, but in many different ways and for many different reasons.
Here are some data-related tasks that a data scientist can handle

V. USE CASES

A. Real Time Analytics by Chicago

The city of Chicago cuts crime and improves citizen welfare with a real-time geospatial analytics platform called windy Grid, which pulls together seven million different pieces of data from city departments every day.
With real-time data becoming more and more available and easily affordable, real time analytics can be what a company may need to push it from just about okay to exceptionally good at understanding themselves, how they are being perceived, their customers wants and needs and much more.
Banks are making out of real time analytics to better engage with their customers.81% of large banks treat customer centricity as their top priority.1 out of every 2 executives believe that they do not have mature capabilities to support their customer strategies.

B. The Solutions To Above Include

1) Don't only act on data from months, weeks or even days ago but also respond to changes that occur second by second.
2) Real time analytics help to understand how to better engage with customer.

How does Technology help?

C. Real Time Data Processing And Analysis

1) Data visualization
2) Predictive Analysis
3) Fraud Detection

Financial Institutions monitor people spending habits on real time basis. Banks and credit card companies collect a lot more information from location, lifestyle, people tastes, income, account balances, employment details, credit history and transaction history. Fraud Instances can be similar in content and appearance but rarely are identical. Data Companies have to constantly update themselves with new technique of fraud detection.

D. Click Stream Analysis

Clickstream data is full trial of digital breadcrumbs left by users as they click their way through a website, and it's loaded with valuable customer information for business. Click stream analysis is process of collecting, analyzing and reporting aggregate data about which pages visitors visit, and in what order .This reveals usage patterns, which in turn gives heightened understanding of customer behavior .This use of the analysis creates a user profile that aids in understanding the types of people that visit a company's website .

E. Sentimental Analysis

Sentimental analysis also known as opinion mining refers to the use of natural processing language. Text Analysis and computational linguistics to identify and extract subjective information in source materials .Using sentimental analysis techniques , companies can respond to negative brand perception. When a

Company releases a new product, monitoring and analyzing social media content can play a large role in quickly remediating bugs and errors.PR for political figures and celebrities depends heavily on sentimental analysis and how person is perceived by people on social media.

F. Analytics for Customer Loyalty Program

Retaining current customers has become a high priority for every business.Loyalty programs have sprung up in every customer related industry, from retail to restaurants, crusine lines to charge cards. Simply a loyalty program cannot differentiate a company from its competitors. Customers expecting discounts and deals wait for special offer sales and use membership cards and simulates a price shopping environment for business rather than increasing their loyalty. Companies that wish to have strong loyalty programs can rely on customer analytics to drive their strategies and create measurable business impact

VI. CONCLUSIONS

This paper gives complete view about Introduction to Data Science and Machine learning standards to deal with the real world problems. This paper mainly concentrated on how data science can be applied to problems in present world. Data Science predict and compute the outcome which would have taken more time for the humans to proceed.

REFERENCES

[1] DataScience.com. “Enterprise Data Science Platfor`m Provider.” DataScience.com, DataScience.com Blog, www.datascience.com/.
[2] DataScience.com. “Data Science Company | Best Data Science Companies.”DataScience.com, DataScience.com Blog, www.datascience.com/company.
[3] “What Is Data Science?” Berkeley University of California, datascience.berkeley.edu/about/what-is-data-science/.
[4] Data Science.” Wikipedia, Wikimedia Foundation, 13 June 2018, en.wikipedia.org/wiki/Data_science.
[5] Data Science.” Coursera, Michigan State University, www.coursera.org/specializations/jhu-data-science.
[6] What Is the DataScience.com Platform?¶.” What Is the DataScience.com Platform? - DataScience.com Platform 4.2.1 Documentation, docs.datascience.com/.
[7]]“DataScience.com.” Crunchbase,
[8] Saltz, Jeffrey S., and Jeffrey M. Stanton. An Introduction to Data Science. SAGE Publications, Inc., 2018. Burlingame, Noreen, and Lars Nielsen. A Simple Introduction to Data Science. New Street Communications, 2012
[9] Nielsen, Lars. A Simple Introduction to Data Science. New Street Communications, 2015.
[10] Igual, Laura, and Santi Segu Ã-. “Introduction to Data Science.” Undergraduate Topics in Computer Science Introduction to Data Science, 2017, pp. 1–4., doi:10.1007/978-3-319-50017-1_1.
[11] Laura, and Santi SeguÃ-. “Toolboxes for Data Scientists.” Undergraduate Topics in Computer Science Introduction to Data Science, 2017, pp. 5–28., doi:10.1007/978-3-319-50017-1_2
[12] Pierson, Lillian, and Jake Porway. Data Science. John Wiley and Sons, Inc., 2017.
[13] Schutt, Rachel, and Cathy O'Neil. Doing Data Science. O'Reilly Media, 2013.
[14] Data Science.” Johns Hopkins Engineering for Professionals, 14 May 2018, ep.jhu.edu/programs-and-courses/programs/data-science.
[15] “Analytics.” IBM Cognitive Advantage Reports, IBM Corporation, www.ibm.com/analytics/data-science.