As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology, 8(05):112-117 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0805018
Available online at: http://www.ijmtst.com/vol8issue05.html

A Survey on Data Science and Its Applications

Dr. Swagata Sarkar[1] | Adarsh V[2] |Adhish S[2] |Divakar K[2]

[1]Head of Department, Artificial Intelligence and Data Science, Sri Sairam Engineering College
[2]Department of Artificial Intelligence and Data Science, Sri Sairam Engineering College
Corresponding Author Email Id: sec20ad001@sairamtap.edu.in

To Cite this Article
Dr. Swagata Sarkar, Adarsh V, Adhish S and Divakar K. A Survey on Data Science and Its Applications. International Journal for Modern Trends in Science and Technology 2022, 8(05), pp. 112-117. https://doi.org/10.46501/IJMTST0805018

Article Info
Received: 30 March 2022; Accepted: 02 May 2022; Published: 04 May 2022.

ABSTRACT

Data science has become an important field in the scientific and research areas. Data science has grown inside various disciplines like statistics, analytics, computing science and also has transformed various domains such as science, engineering, social science, etc. This area includes big data, machine learning, artificial intelligence, natural language processing, etc. The scope of data science is large enough in the present as well as future considering its impact in the practical world.

Keywords-Data Science, Artificial Intelligence,Machine learning, Natural Language Processing, Big Data.

1. INTRODUCTION

The field Data science composes of aggregating, analysing, cleansing, manipulating the data to perform advanced data analysis. Various machine learning algorithms are used in data science to build predictive models. These processes are achieved through statistics, scientific and AI methods. This analytical application can be reviewed the results to uncover the patterns. People who work on data science are called data scientists [1].

Data science is very crucial for business needs. One of the common challenges faced by many companies is to analyse and categorize the given data. Hence knowing how to handle the bunch of data properly and efficiently, a company or organization can progress a lot [2].

Everyday a lot of data is generated as time goes this data becomes large and complex, hence the scope of data science is undoubtedly large.

Data Science has various application in various fields like Healthcare, Fraud Detection, Robotics, Finance, E-Commerce, Transportation, Image Processing, Recommendation engines

The main application of data science from olden days is in finance. Banking companies use data science for risk and fraud detection [7].

Many of the of the advertisements we see on the television are based on the data collected on audience and it is shown for targeted audience.

Almost all of the searches we make are driven by data science in today's world [6]. It turns and allows many companies to costs management, increase efficiencies, identify new opportunities in market, and it boost their advantage and personal assistant like Siri for data science recommendation demands. Customers service are being handled by chatbots now a days . Other areas where data science play a major role are image recognition, speech recognition, gaming, flight management etc. [3].

2. EMERGENCE OF DATA SCIENCE

It was in early 1960s where the term data science was first introduced to introduce a new profession which would handle large amount of data. Data science is evolving as a discipline using computer science and statistical method to find the new predictions. Data science has evolved deeply to use Artificial intelligence, machine learning and IoT. With the development of internet, the concept of big data has become a trend now. In 1962 John Turkey in his paper titled The Future of Data Analysis, described the shift in statistics. In his second paper in 1977 he stated the importance of data analysis in data science. Data science has grown in large rate in past 30 years. It is now used by governments, geneticist, engineers, etc world-wide. A shift in use of bigdata in data science is considered to increase its growth rapidly in coming years too. Figure 1 and figure 2 shows about the job growth and evolution of data science respectively [4].

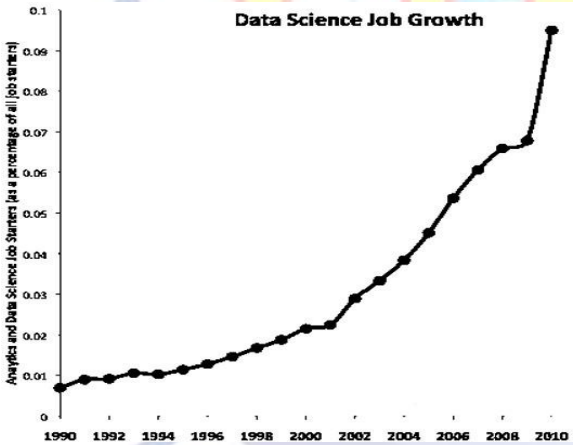

Figure 1 Growth of Data Science

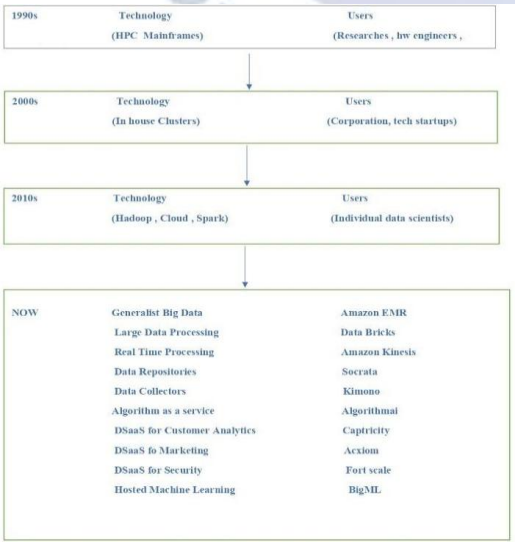

Figure 2 Emergence of Data Science

3. DATA SCIENCE TOOLS

Data science is a most popular filed in 21st century and its growth is drastically large as shown in figure 1. A Data scientist is highly responsible in analysing and managing huge quantity of structured and unstructured data. To achieve this, he requires large number of tools and programming languages for working on data science. Some of the data science tools has been covered in this paper.

Here are some of tools in data science used by most of the data scientists till now.

I) SAS: SAS is one of the most used data science tools used mainly for statistical operations. SAS is not an open-source software but used by most of the large organisations or performing statistical modelling.
This software provides numerous statistical libraries and tools for data scientist. even though SAS is highly reliable it is very expensive and is being used by larger companies only [5].

II) BigML: BigML is another tool that is commonly used by data scientists. It offers GUI that is cloud based which can be used for processing Machine Learning Algorithms. It is specialized in predictive modelling. BigML uses huge types of algorithms like classification, clustering, time series forecasting [5].

III) *MATLAB:* MATLAB is a multi-paradigm numerical software for processing mathematical facts. It is also not anopen source software.IT allows statistical modelling of data. It is used for simulation of neural networks and fuzzy logic. It also enables processing of signals and image. It provides various visualisation tools [5].

IV) *Excel:* It is probably most used data analytics tool. It wasdeveloped by Microsoft for spread sheet calculations, today it is widely used for data processing, visualisation and complex calculations. It allows usage of various formulae, tables. filters, slicers, etc. It is also an excellent data cleaning tool [5].

V) *ggplot2:* This is a common R programming visualization software. It is the most commonly used library for generating visualizations from analysed data. It is even possible to annotate the data in visualisations, add text labels [5].

VI) Tableau: Tableau is a software that has interactive visualization. The most fascinating side of tableau is its functionality to link with spreadsheets, OLAP, databases, etc. Tableau has large community in online. The free version of Tableau is called Tableau Public [5].

VII) TensorFlow: It is a common tool in data science. Widely used software for advanced machine learning algorithms like deep learning. It is an open-source software that has been evolving since its created. It can operate on both CPUs and GPUs. It is one of the best in terms of advanced machine learning algorithm processing. Some of the applications of TensorFlow are speech recognition, image classification and language generations, etc [5].

VIII) Matplotlib: Python has Developed a library called Matplotlib which is very useful for plotting and visualizing the data. It is very popular for creating graphs with the data's provided. Using simple lines of code, a user can create or plot the any complicated graphs for the given data. One of the main modules in Matplotlib is Pyplot and is used by Data Analysts to visualize the data. It is an open- source software hence can be used by any one and a very good tool for beginners to learn data visualization [5]

4. DATA SCIENCE APPLICATIONS

There are various Data science Applications revolving around us some the applications are:

A. Healthcare

This sector receives highest benefits from data science applications [3]

I) Medical Image Analysis: Frameworks like MapReduce are used to find optimal parameters for tasks like lung texture Classification. It uses machine learning methods, support vector machines and various other classification techniques too.

II) Developing Drugs: The process of drug discovery is highly complicated thus data science applications and machine learning algorithms simplify this process. The central idea of drug discovery is to generate computer model simulations as biologically relevant network for simplification of prediction of future outcomes with high accuracy

III) Patient Assistance: The concept of travel less patient check-up is becoming possible through applications of data science.AI powered mobile apps usually containing chatbots provide health care support. The data collected from patients help them to lead a better life and it also helps the doctors to suggest more effective medicines.

B. Internet Search

This is the most common topic that strikes everybody's mind when we talk of data science application. All the search engines make use of data science algorithm to provide best results for our search. Various user data is being collected from us on every search we make This is being effectively used to provide bet user experience in search engines [3].

C. Targeted Advertising

Past years has seen big growth in advertising. Advertising through online gives faster feedback and publishers more knowledge about their customers. The concept of spending small amount of money and obtaining maximum profit can be obtained through targeted advertising. This involves when, where and to whom display particular advertisement. The main work of machine learning system is to identify the customers who will be able to buy a particular product in coming future after being displayed the advertisement. A primary source of such algorithm is user browsing history that is websites that is visited in the past.

The most common used approach in data science is to use proxy trained models. Earlier click thorough rate was considered as important criteria for targeted advertising. But having higher click through rate doesn't mean higher customers will be obtained in the future [6].

A. Augmented Reality

Augmented Reality, it is our reality which is augmented with digital data. Pictures, texts, videos, 3d assets, or a combination these are digital data.AR should understand the reality and reconstruct to create its own digital twin.AR need to enable the user to interact with both the digital data and digital twin. Cameras didn't have any inherent sense of direction. No one can command it to take pictures of objects. You can only point it in the direction that you want. It is an array of sensors that burn up when light rays shooting through a hole pour down on it. Understanding of the world and the camera a sense of direction is a complex process. The study of field which deals with understanding and

analysing pixels is called Computer Vision. Machine learning, Artificial Intelligence and Computer Vision including some hardware sensors that form the Inertial Measurement Unit, make Augmented Reality possible [8].

B. Fraud Detection

Fraud happens in every industry you go or see; it affects everyone including the industry itself although not equally but somewhat. To identify why and when the frauds happen the industry which face these problems use data analysis to help finding out those above-mentioned question.

I)Data Analytics is used in finding the pharmaceutical fraud: Fraud in healthcare sector can happen when a provider prescribes a drug to a patient who doesn't actually have the need of it or if the medical store charges high prices for a medicine if the that medicine has high demand in the market. These are some fraud cases which are common and prevailing in the society.

Data Analytics could help in such cases like if a patient has some sort of disease and is taking same medication for a long interval, when suddenly medication changes a user can check why there is such change in medication. One more such case is in detecting cases of pharmacy refill fraud as pharmacy refills a prescription medication before itself even patient has not request it. We can use data analytics in checking the timeline of generic drugs and whohas requested it like that.

Data Analytics is also very useful in detecting fraud in financial or banking sector we can check a pattern of fraud happened around the country and can make a predictive machine learning model and feeding those data into it, then using this model we can predict when the fraud can happen in the future. Using these models, we can have 24/7 surveillance and can intimate if any kind of disturbance is observed [10].

D. Recommendation Systems

Recommendation systems shows us the most popular contents or searches. Recommendation systems are based on three factors:

1.Users
2.Content
3.Rating

Two different approaches for creating recommendation systems are:

I) Content-based: The content-based approach is completely based on user data. Users' data or content is used to target a new user that falls under the same category.

II) Collaborative Filtering: Collaborative filtering is complex when compared to content-based method. This method is based on the ratings given by users. The main purpose of the collaborative filtering is to predict the ratings for each subject and each user. Collaborative filtering is more accurate when compared to content-based filtering as the model works on large datasets with more power. Hence recommendation systems are one the most widely used application in data science[9].

E. Gaming

Gaming has been exploding in recent years. The number of active users grows for every minute, as does the overall revenue of game development industry. The games' internal infrastructure becomes more and more complex, giving user more options. For the users, an entirely new reality is created. Users are highly satisfied thanks to advanced Graphics and techniques, as well as the latest visual effects and augmented reality effects.

Data science has improved the principles of operation in a variety of company, propelled a different type of businesses to a higher level of development and it is true in the gaming industry as well. Lot of data science techniques and methodologies have become essential in thedevelopment and other operations of games. Finding transferable algorithms by creating artificial intelligence systems. Data scientist analyse data from in-depth learning games. This learning will be appealing for data scientist's view in finding common ways to use in project in sports and also in future programs and other games as well. Data scientist investigates, designs and uses ideas, and organizes

different variety of ideas. They are also responsible for creating automated analytics tools and mathematical models for identifying. Mathematics team, experienced data scientist should perform large data analytics in large data and develop, predictive and descriptive models using depth data learning algorithms.

5. DATA SCIENCE ADVANTAGES

The data science field has its own advantages and disadvantages. So here we can discuss the pros and cons of data science.

A. High Demand

Data science is in huge demand. It has been rapidly growing job on social medias and it is predicted to create

11.5 million jobs by 2026.This makes it high demanding job in present years [11].

B. It is versatile

There are various applications for data science. It is very much used in healthcare, finance, consulting, e-commerce and other sectors. Therefore, a data scientist has various filed to work on [11].

C. Highly paid

Data science is a high paid job. On an average a data scientist earns $116,100 per year. This makes data science so popular [11].

D. Improves Data quality

A data scientist should also improve the quality of data. Their Data science deals with processing, analysing and improving the quality of the data [11].

E. Makes smarter product

With the use of machine learning data science can create better predictions. With the growth of data science, it has made computers to understand human behaviours and take decision based on data [11].

F. Can save lives

Data Science has improved health care sector a lot. It is now possible to detect early-stage tumours. Other medical diagnostics has also been improved with the help of data science thus it has saved many lives [11].

6. DISADVANTAGES

A. Mastering is impossible

Data science depends on mathematics, statistics and computer science. Therefore, it is not possible to master each field and becoming expert in this area. This is an ever-changing field [11].

B. Lack of domain knowledge

Data science largely depends on domain knowledge. A person with less experience in statistics and computer science will not be able to achieve in the field of data science. One must have large experience in this domain to achieve in data science. This can increase the difficulty in changing from one domain to another [11].

C. Data privacy

A data scientist needs to handle large amount of user data. This data may be of user's sensitive information like citizen Id etc. Hence there is always a risk in protecting these data from leaking to others. This raises many ethical issues [11].

D. Blurry Term

The data science is a blurry term and does not contain accurate definition. Role of a data scientist depends on thearea he/she is working on[11].

7. CONCLUSION

Data science has significantly improved the standard of living in almost every possible way. It has almost become tough to imagine a domain without the influence of data science. Day by day as the size of the data increases the need for data science is also increasing. Still there is a huge requirement for quality data scientist with good ethical values. It is highly predicted that data science will improve the quality of various domains. This paper presented the various applications, advantages, growth of data science.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Data Science definition [online] https://bit.ly/3N7XNiE

[2] Future Scope of Data Science [online]https://www.edureka.co/blog/future-scope-of-data-science/

[3] Data Science applications [online] https://www.edureka.co/blog/data- science-applications/

[4] Data science evolution [online] https://www.dataversity.net/brief-history-data-science/#

[5] Data Science tools [online] https://data-flair.training/blogs/data-science-tools/

[6] Targeted Advertising [online]https://labs.eleks.com/2014/02/data-science-for-targeted-advertising.html

[7] fraud detection https://www.henryharvin.com/blog/usage-of-data-science-in-fraud-detection/

[8] Augmented Reality [online] https://www.edureka.co/blog/data-science-applications/#augreality

[9] Recommendation Systems [online] https://thecleverprogrammer.com/2021/10/14/recommendation-systems- in-data-science/
[10] Fraud Detection [online] https://www.henryharvin.com/blog/usage-of-data-science-in-fraud-detection/
[11] Advantages and Disadvantages of Data Science [online] https://data- flair.training/blogs/pros-and-cons-of-data-science/