Introduction of Data Science

Presentation · July 2020

CITATIONS	READS
0	7,032

1 author:

Nilu Singh
Koneru Lakshmaiah Education Foundation
120 PUBLICATIONS **366** CITATIONS

Introduction of Data Science

Dr. Nilu Singh

School of Computer Applications

Babu Banarasi Das University

Lucknow-UP

Content

- Introduction of Data Science
- Need of Data Science
- Data Science Components
- Tools for Data Science
- Applications of Data Science
- Summary
- References

Introduction

- *The simplest definition of data science is the extraction of actionable insights from raw data.*

- Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge.
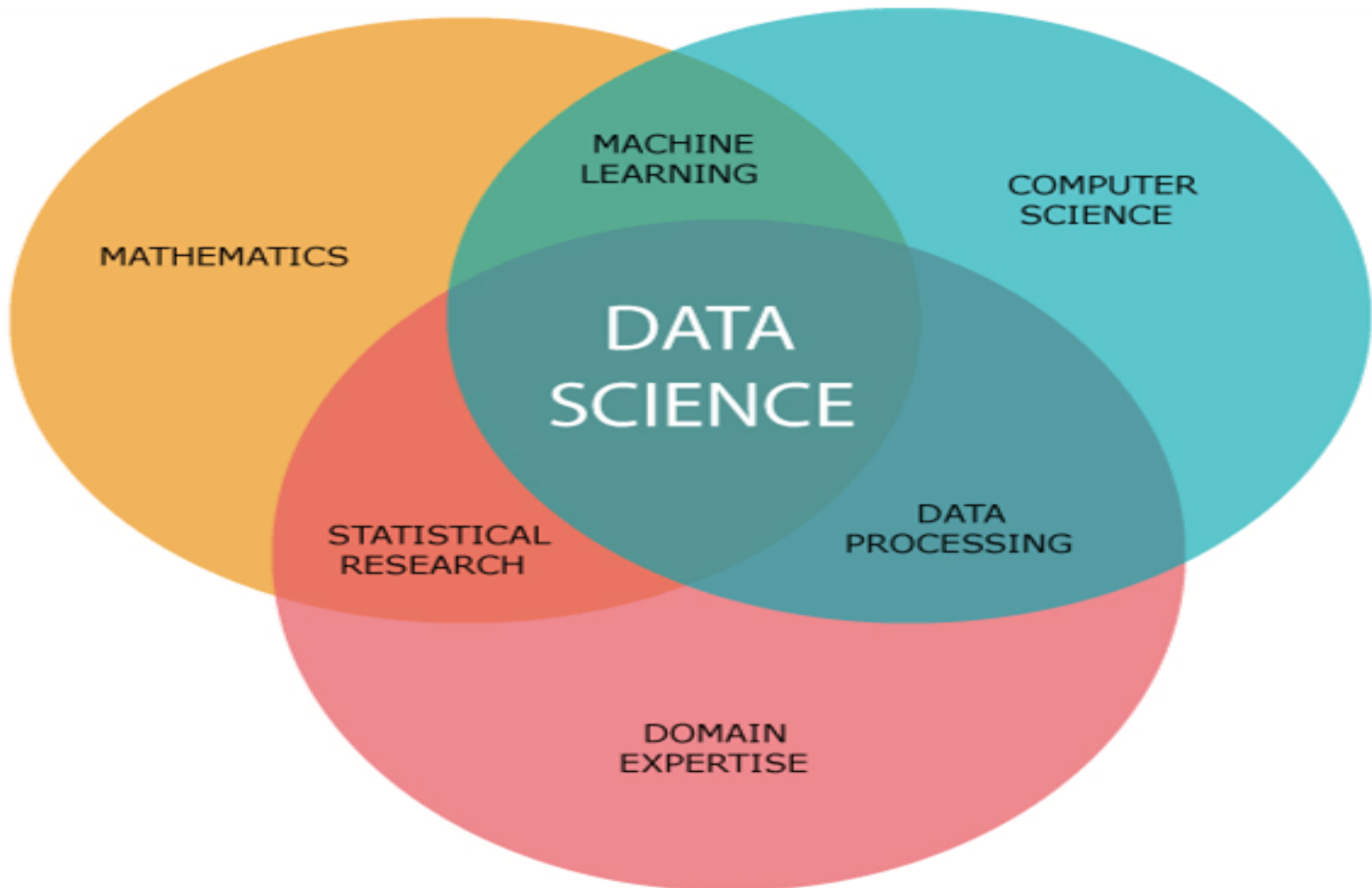
Conti…

- Data science is a deep study of the massive amount of data, which involves extracting meaningful insights from raw, structured, and unstructured data that is processed using the scientific method, different technologies, and algorithms.

- It is a multidisciplinary field that uses tools and techniques to manipulate the data so that you can find something new and meaningful.

- Data science uses the most powerful hardware, programming systems, and most efficient algorithms to solve the data related problems. It is the future of artificial intelligence.
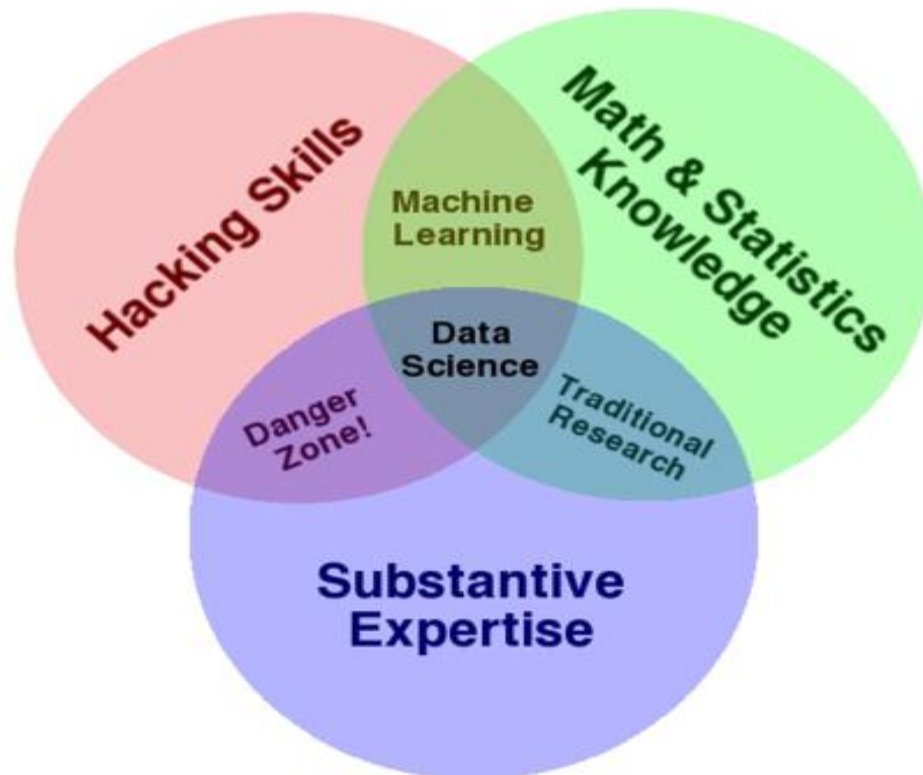
Conti…

- **Data science** is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data.

- Data science is related to data mining, deep learning and big data.

- It is a "concept to unify statistics, data analysis, machine learning, domain knowledge and their related methods" in order to "understand and analyze actual phenomena" with data.

Conti…

Conti…

Data Science as a field focused on extracting knowledge and insights from data by using scientific methods.

Conti…

- It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, domain knowledge and information science.

- An area that manages, manipulates, extracts, and interprets knowledge from tremendous amount of data.

- Data science (DS) is a multidisciplinary field of study with goal to address the challenges in big data.

- Data science principles apply to all data i.e. big and small.

History

- The term "data science" has been traced back to 1974, when Peter Naur proposed it as an alternative name for computer science.

- In 1996, the International Federation of Classification Societies became the first conference to specifically feature data science as a topic.

- In 1997, C.F. Jeff Wu suggested that statistics should be renamed data science

- in 2002, the Committee on Data for Science and Technology launched Data Science Journal.

Conti…

- In 2003, Columbia University launched The Journal of Data Science.

- In 2014, the American Statistical Association's Section on Statistical Learning and Data Mining changed its name to the Section on Statistical Learning and Data Science, reflecting the ascendant popularity of data science

# Uses of Data Science

we can say that data science is all about:

•Asking the correct questions and analyzing the raw data.

•Modeling the data using various complex and efficient algorithms.

•Visualizing the data to get a better perspective.

•Understanding the data to make better decisions and finding the final result.

Conti…

Example:

Let suppose we want to travel from station A to station B by car. Now, we need to take some decisions such as which route will be the best route to reach faster at the location, in which route there will be no traffic jam, and which will be cost-effective. All these decision factors will act as input data, and we will get an appropriate answer from these decisions, so this analysis of data is called the data analysis, which is a part of data science.

Need for Data Science:
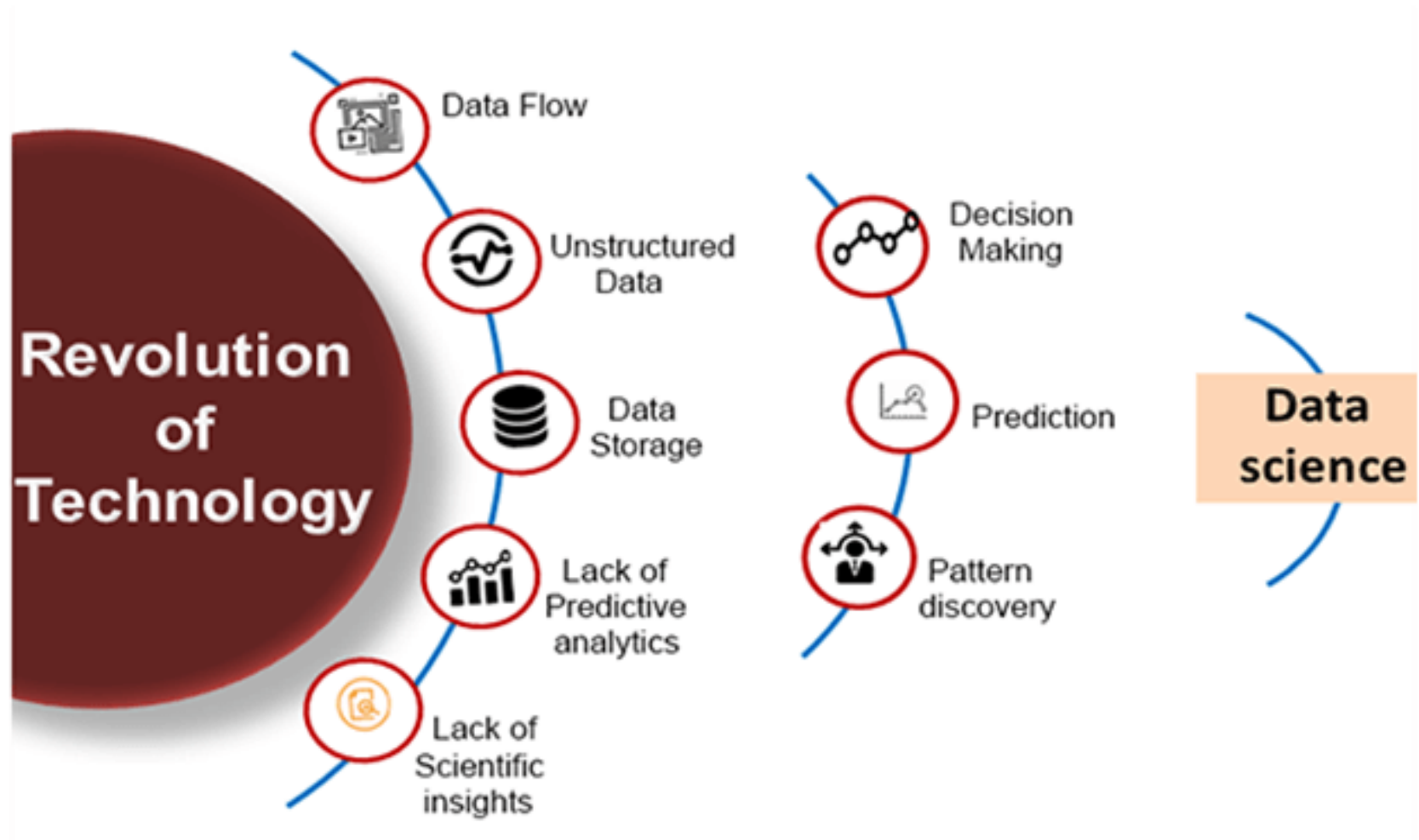
Conti…

some years ago, data was less and mostly available in a structured form, which could be easily stored in excel sheets, and processed using BI (business intelligence) tools.

•But in today's world, data is becoming so vast, i.e., approximately 2.5 quintals bytes of data is generating on every day, which led to data explosion.

•It is estimated as per researches, that by 2020, 1.7 MB of data will be created at every single second, by a single person on earth. Every Company requires data to work, grow, and improve their businesses.
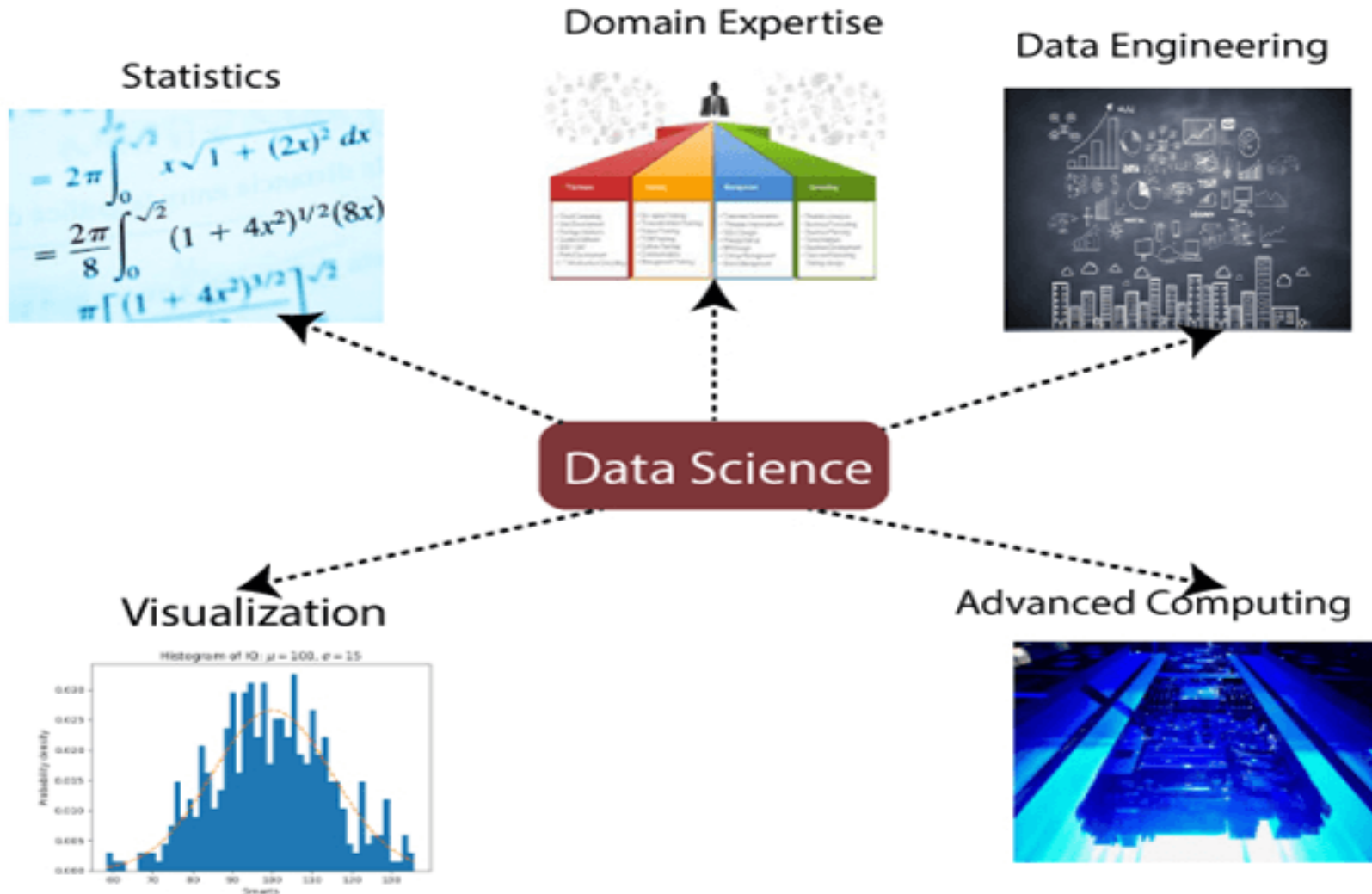
Conti…

- It is estimated as per researches, that by 2020, 1.7 MB of data will be created at every single second, by a single person on earth. Every Company requires data to work, grow, and improve their businesses.

- Now, handling of such huge amount of data is a challenging task for every organization. So to handle, process, and analysis of this, we required some complex, powerful, and efficient algorithms and technology, and that technology came into existence as data Science.

Conti…

Following are some main reasons for using data science technology-

•With the help of data science technology, we can convert the massive amount of raw and unstructured data into meaningful insights.

•Data science technology is opting by various companies, whether it is a big brand or a startup. Google, Amazon, Netflix, etc, which handle the huge amount of data, are using data science algorithms for better customer experience.

•Data science is working for automating transportation such as creating a self-driving car, which is the future of transportation.

•Data science can help in different predictions such as various survey, elections, flight ticket confirmation, etc.

Data Science Components

Statistics

Domain Expertise

Data Engineering

Data Science

Visualization

Advanced Computing

Conti…

The main components of Data Science are given below:

1. **Statistics:** Statistics is one of the most important components of data science. Statistics is a way to collect and analyze the numerical data in a large amount and finding meaningful insights from it.

2. **Domain Expertise:** In data science, domain expertise binds data science together. Domain expertise means specialized knowledge or skills of a particular area. In data science, there are various areas for which we need domain experts.

3. **Data engineering:** Data engineering is a part of data science, which involves acquiring, storing, retrieving, and transforming the data. Data engineering also includes metadata (data about data) to the data.

4. **Visualization:** Data visualization is meant by representing data in a visual context so that people can easily understand the significance of data. Data visualization makes it easy to access the huge amount of data in visuals.
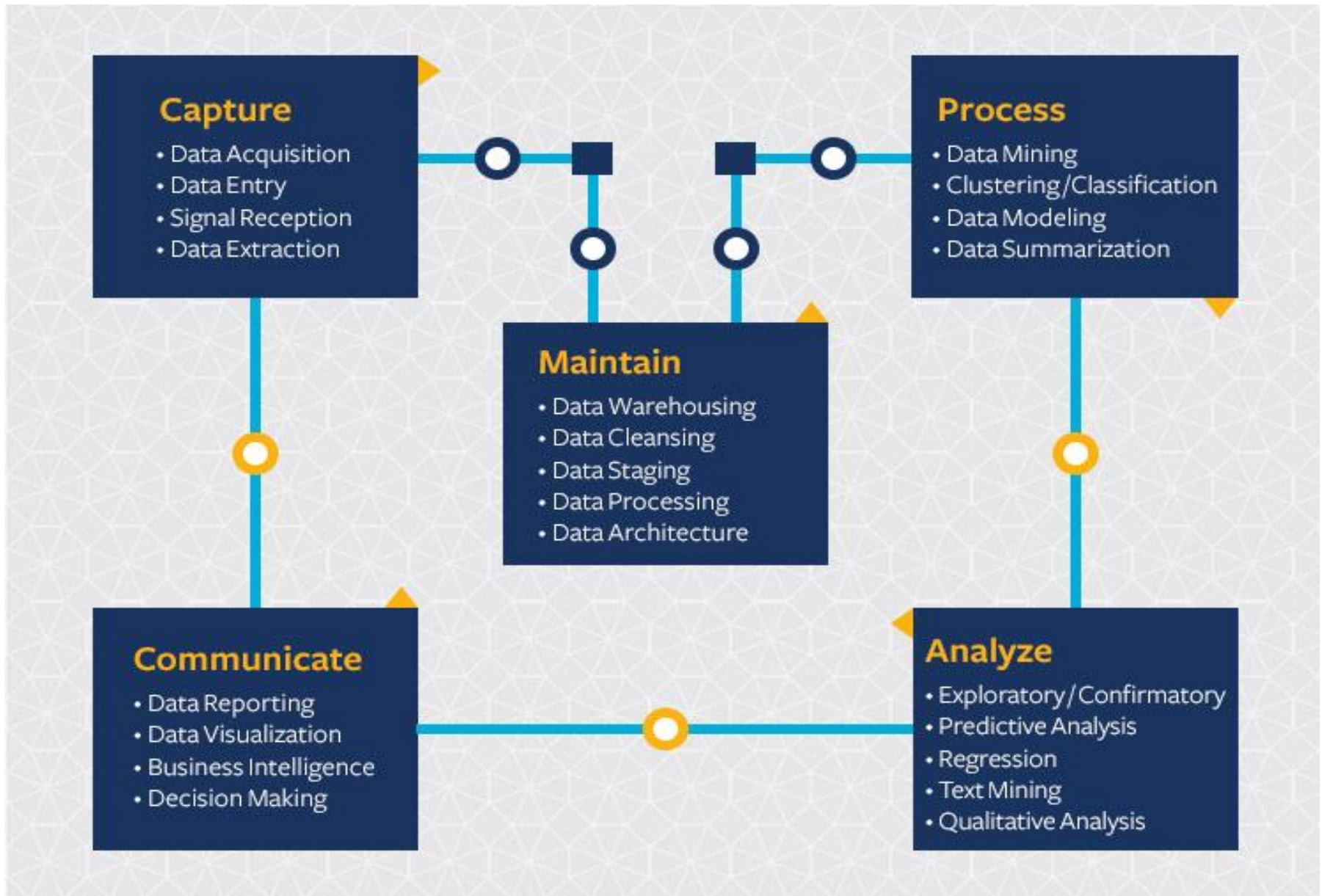
5. **Advanced computing:** Heavy lifting of data science is advanced computing. Advanced computing involves designing, writing, debugging, and maintaining the source code of computer programs.

6. **Mathematics:** Mathematics is the critical part of data science. Mathematics involves the study of quantity, structure, space, and changes. For a data scientist, knowledge of good mathematics is essential.

7. **Machine learning:** Machine learning is backbone of data science. Machine learning is all about to provide training to a machine so that it can act as a human brain. In data science, we use various machine learning algorithms to solve the problems.

Data Science Life Cycle

Capture
- Data Acquisition
- Data Entry
- Signal Reception
- Data Extraction

Process
- Data Mining
- Clustering/Classification
- Data Modeling
- Data Summarization

Maintain
- Data Warehousing
- Data Cleansing
- Data Staging
- Data Processing
- Data Architecture

Communicate
- Data Reporting
- Data Visualization
- Business Intelligence
- Decision Making

Analyze
- Exploratory/Confirmatory
- Predictive Analysis
- Regression
- Text Mining
- Qualitative Analysis

Conti…

The explanation of the data science life cycle is-

Capture: data acquisition, data entry, signal reception, data extraction).

Maintain: data warehousing, data cleansing, data staging, data processing, data architecture).

Process: data mining, clustering/classification, data modeling, data summarization.

Analyze: exploratory/confirmatory, predictive analysis, regression, text mining, qualitative analysis).

Communicate: data reporting, data visualization, business intelligence, decision making.

Impacts of data science

- Big data is very quickly becoming a vital tool for businesses and companies of all sizes.
- The availability and interpretation of big data has altered the business models of old industries and enabled the creation of new ones.
- Data scientists are responsible for breaking down big data into usable information and creating software and algorithms that help companies and organizations determine optimal operations.
- As big data continues to have a major impact on the world, data science does as well due to the close relationship between the two.

Conti…
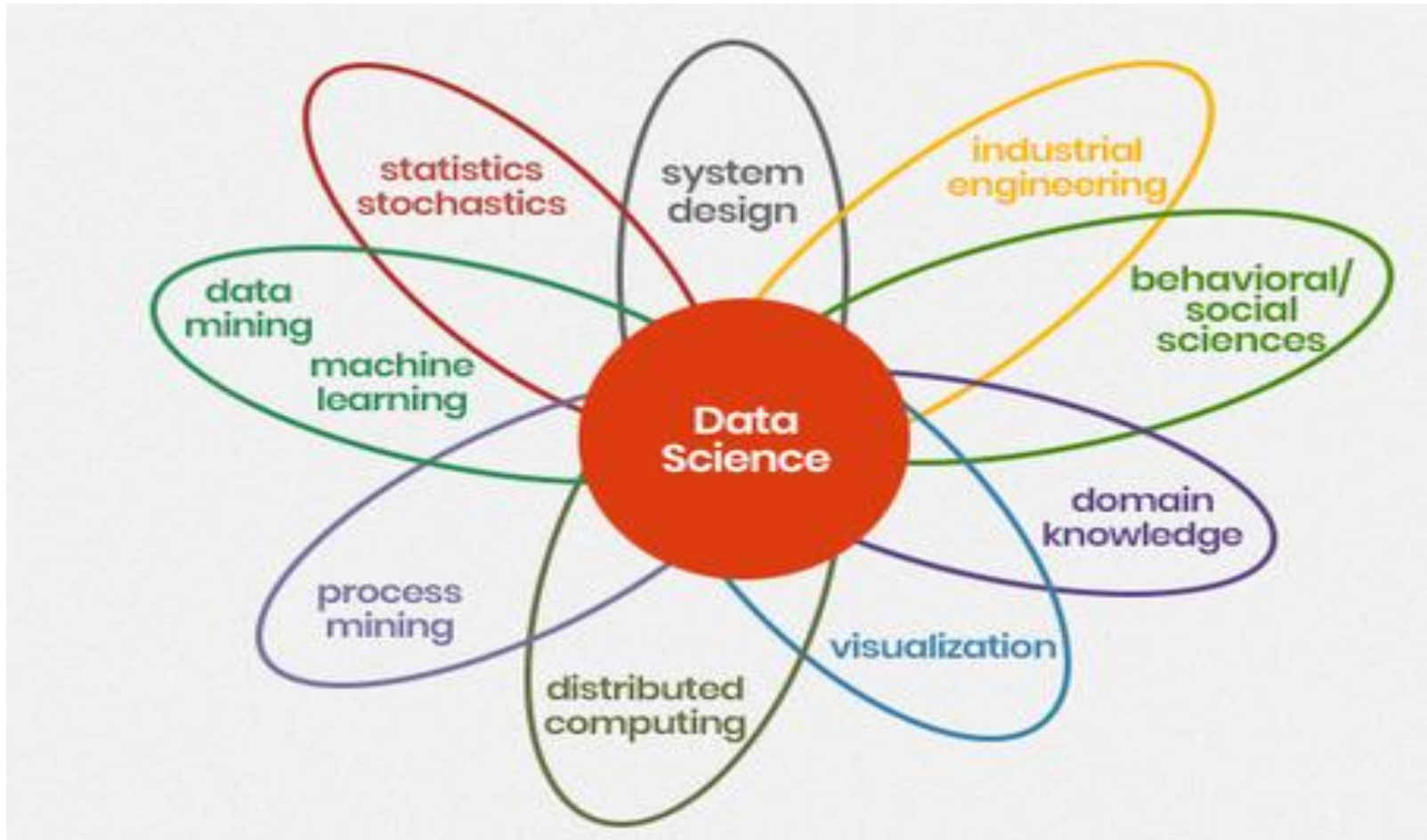
Data Science will Impact Future of Businesses

Data Science and Prediction

- Data science might therefore imply a focus involving data and, by extension, statistics

- or the systematic study of the organization, properties, and analysis of data and its role in inference, including our confidence in the inference.

Ques: Why then do we need a new term like data science when we have had statistics for centuries?

Ans: The fact that we now have huge amounts of data should not in and of itself justify the need for a new term.

Tools for Data Science

Following are some tools required for data science:

Data Analysis tools: R, Python, Statistics, SAS, Jupyter, R Studio, MATLAB, Excel, RapidMiner.

Data Warehousing: ETL, SQL, Hadoop, Informatica/Talend, AWS Redshift

Data Visualization tools: R, Jupyter, Tableau, Cognos.

Machine learning tools: Spark, Mahout, Azure ML studio.

Applications of Data Science

There are some common application areas of Data Science-

➢ Fraud and Risk Detection
➢ Healthcare
➢ Internet Search
➢ Targeted Advertising
➢ Website Recommendations
➢ Advanced Image Recognition
➢ Speech Recognition
➢ Airline Route Planning
➢ Gaming
➢ Augmented Reality

Conti…

- **Anomaly detection** (fraud, disease, crime, etc.)
- **Automation and decision-making** (background checks, credit worthiness, etc.)
- **Classifications** (in an email server, this could mean classifying emails as "important" or "junk")
- **Forecasting** (sales, revenue and customer retention)
- **Pattern detection** (weather patterns, financial market patterns, etc.)
- **Recognition** (facial, voice, text, etc.)
- **Recommendations** (based on learned preferences, recommendation engines can refer you to movies, restaurants and books you may like) and many more.

Conti…

Image recognition and speech recognition:
Data science is currently using for Image and speech recognition. When you upload an image on Facebook and start getting the suggestion to tag to your friends. This automatic tagging suggestion uses image recognition algorithm, which is part of data science.
When you say something using, "Ok Google, Siri, Cortana", etc., and these devices respond as per voice control, so this is possible with speech recognition algorithm.

Gaming world:
In the gaming world, the use of Machine learning algorithms is increasing day by day. EA Sports, Sony, Nintendo, are widely using data science for enhancing user experience.

Internet search:
When we want to search for something on the internet, then we use different types of search engines such as Google, Yahoo, Bing, Ask, etc. All these search engines use the data science technology to make the search experience better, and you can get a search result with a fraction of seconds.

Conti…

Transport:
Transport industries also using data science technology to create self-driving cars. With self-driving cars, it will be easy to reduce the number of road accidents.

Healthcare:
In the healthcare sector, data science is providing lots of benefits. Data science is being used for tumor detection, drug discovery, medical image analysis, virtual medical bots, etc.

Recommendation systems:
Most of the companies, such as Amazon, Netflix, Google Play, etc., are using data science technology for making a better user experience with personalized recommendations. Such as, when you search for something on Amazon, and you started getting suggestions for similar products, so this is because of data science technology.

Risk detection:
Finance industries always had an issue of fraud and risk of losses, but with the help of data science, this can be rescued.
Most of the finance companies are looking for the data scientist to avoid risk and any type of losses with an increase in customer satisfaction.

Machine learning in Data Science

To become a data scientist, one should also be aware of machine learning and its algorithms, as in data science, there are various machine learning algorithms which are broadly being used. Following are the name of some machine learning algorithms used in data science:

- Regression
- Decision tree
- Clustering
- Principal component analysis
- Support vector machines
- Naive Bayes
- Artificial neural network
- Apriori

Summary

- Big Data has given rise to Data Science.
- Data science is rooted in solid foundations of mathematics and statistics, computer science, and domain knowledge.
- New profession –Data Scientists, Data Engineer (DE), Data Analyst (DA), Application architect (AA)
- Additionally, how businesses are using data science to innovate in their sectors.
- The term "data scientist" was coined as recently as 2008 when companies realized the need for data professionals who are skilled in organizing and analyzing massive amounts of data.

Types of Data Science Job

The average salary range for data scientist will be approximately $95,000 to $ 165,000 per annum, and as per different researches, about 11.5 millions of job will be created by the year 2026.

- The main job roles are given below:
- Data Scientist
- Data Analyst
- Machine learning expert
- Data engineer
- Data Architect
- Data Administrator
- Business Analyst
- Business Intelligence Manager

References

Thank You!

Questions?