DATA SCIENCE

Data Science is the area of study involving the extraction of information from vast amounts of data through the use of different scientific methods, algorithms, and processes. It helps you find hidden patterns out of the raw data. Owing to the emergence of mathematical statistics, data mining, and big data, the word data science has arisen. It allows you to extract knowledge from both structured and unstructured data. Data science helps you to convert a business question into a research project and then return it to a practical solution. There are several components in data science they are as follows:

Statistics: Throughout Data science, data is the most important unit. It is the technique to collect large quantities of numerical data and analyze to gain useful information.

Visualization: The visualization methodology lets you view huge amounts of data in images that are easy to understand and digestible. There are many approaches to data visualization, The most popular is the presentation of details, usually integrating statistical graphics and thematic cartography. There are several areas of concentration they are:

- Displaying news
- Displaying data
- Displaying connections
- Displaying websites etc

Machine learning: machine learning is a scientific study of algorithms and statistical models used by computer systems to perform a particular task without using explicit instructions. It builds a mathematical data based on sample data known as "training data" in order to make predictions or decisions without being explicitly programmed.

Deep learning: Deep Learning method is new machine learning research where the algorithm selects the analysis model to follow.

Data science process: Manager

1. **Discovery:** Acquiringdata from different sources like internal and external which helps you to answer different questions. The data can be from social networking sites and webservers etc.
2. **Data preparation:** After the discovery of data we should prepare the data. Data can have lots of inconsistencies like missing value, blank columns, incorrect data format and some other problems which need to be cleaned.
3. **Model planning:** After the data planning we should plan about the model. In this stage, we should plan the model technique and relationships among them. Planning for a model

involves a few steps such as statistical formulas and visualization and R are some tools used for model planning.

4. **Model building:** After the planning model we should build a model. The actual model-building process starts in this stage. Here, data scientist distributes the training and testing datasets. The training data set is applied to methods such as correlation, classification, and clustering. Once prepared the model will be tested against the dataset "testing." Both testing and training plays an important role.

5. **Operationalize:** You provide the final baselined model with reports, code, and technical documents at this level. Upon thorough testing, the model is implemented into a production environment in real-time.

6. **Communicate Results:** In this stage, the key findings are communicated to all stakeholders. This helps you to decide if the results of the project are a success or a failure based on the inputs from the model.

Tools used in data science: R language, python, SQL, Java, MATLAB

Applications of Data science:

Internet Search: Google search use Data science technology to search a specific result within a fraction of a second

Recommendation Systems: To create a recommendation system. Example, "suggested friends" on Facebook or suggested videos" on YouTube, everything is done with the help of Data Science.

Speech Recognition: Speech recognizes systems like Siri, Google Assistant, Alexa runs on the technique of Data science. Moreover, Facebook recognizes your friend when you upload a photo with them, with the help of Data Science.

Gaming world: EA Sports, Sony is using Data science technology. This enhances your gaming experience. Games are now developed using Machine Learning using data science.

Why Data Scientists are called 'Data Scientists'?

Upon considering the fact that a data scientist gathers a huge amount of information from the scientific fields and applications whether the information is statistical, mathematical or computer science, the word "Data scientist" has been in existence. We use the latest technologies and resources to find solutions and arrive at the conclusions that are important to an organization's growth and development. Data Science has gained vast inference and has extended its roots deep in the IT industry. This popularity has made students and professionals to master data science. To curb this, various Data Science courses are available through which one can become Data Scientist with no difficulty.

Data Science Jobs Roles:

Most prominent Data Scientist job titles are:

- Data Scientist
- Data Engineer
- Data Analyst
- Statistician
- Data Architect
- Data Admin
- Business Analyst
- Data/Analytics

Data Scientist:

Role: A Data Scientist is a professional who manages enormous amounts of data to come up with compelling business visions by using various tools, techniques, methodologies, algorithms, etc.
Languages: R, SAS, Python, SQL, Hive, Matlab, Pig, Spark

Data Engineer:

Role: The role of a data engineer is of working with large amounts of data. He develops, constructs, tests, and maintains architectures like large scale processing systems and databases.
Languages: SQL, Hive, R, SAS, Matlab, Python, Java, Ruby, C + +, and Perl

Data Analyst:

Role: A data analyst is responsible for mining vast amounts of data. He or she will look for relationships, patterns, trends in data. Later he or she will deliver compelling reporting and visualization for analyzing the data to take the most viable business decisions.
Languages:[R programming](#), Python, HTML, JS, C, C+ + , SQL

Statistician:

Role: The statistician collects, analyses, understand qualitative and quantitative data by using statistical theories and methods.
Languages: SQL, R, Matlab, Tableau, Python, Perl, Spark, and Hive

Data Administrator:

Role: Data admin should ensure that the database is accessible to all relevant users. He also makes sure that it is performing correctly and is being kept safe from hacking.
Languages: Ruby on Rails, SQL, Java, C#, and Python

Business Analyst:

Role: This professional needs to improves business processes. He/she as an intermediary between the business executive team and the IT department.
Languages: SQL, Tableau, Power BI and, Python
Challenges in data science:

- Data Science results not effectively used by business decision-makers
- Explaining data science to others is difficult
- Privacy issues
- Lack of significant domain expert
- If an organization is very small, they can't have a Data Science team