1. 1 INTRODUCTION :

What is Data Science?

Data science continues to evolve as one of the most promising and in-demand career paths for skilled professionals. Today, successful data professionals understand that they must advance past the traditional skills of analyzing large amounts of data, data mining, and programming skills.

In order to uncover useful intelligence for their organizations, data scientists must master the full spectrum of the data science life cycle and possess a level of flexibility and understanding to maximize returns at each phase of the process.

Data science is all about uncovering findings from data. Diving in at a granular level to mine and understand complex behaviors, trends, and inferences. It's about surfacing hidden insight that can help enable companies to make smarter business decisions.

For example:

- ➔ Netflix data mines movie viewing patterns to understand what drives user interest, and uses that to make decisions on which Netflix original series to produce.
- ➔ Target identifies what are major customer segments within its base and the unique shopping behaviors within those segments, which helps to guide messaging to different market audiences.
- ➔ Proctor & Gamble utilizes time series models to more clearly understand future demand, which help plan for production levels more optimally.

Applications of Data Science:

➢ Fraud and Risk Detection

The earliest applications of data science were in Finance. Companies were fed up of bad debts and losses every year. However, they had a lot of data which use to get collected during the initial paperwork while sanctioning loans. They decided to bring in data scientists in order to rescue them out of losses.

➢ Targeted Advertising

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 airports – 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 behaviour.

➢ Advanced Image Recognition

You upload your image with friends on Facebook and you start getting suggestions to tag your friends. This automatic tag suggestion feature uses face recognition algorithm.

Difference between Data Science, Data Analytics and Big Data…

Big Data	Data Science	Data Analytics
• Retail • Banking and investment • Fraud detection and analyzing • Customer-centric applications • Operational analysis	• Web development • Digital advertisements • E-commerce • Internet search • Finance • Telecom • Utilities	• Travelling and transportation • Financial analysis • Retail • Research • Energy management • Healthcare
	Skills Required	
• Analytical skills • Mathematics and statistics • Java • Hadoop	• SAS • R/Python programming • Hadoop • SQL database • Analytical skills • Statistics • Mathematics • Visionary thinking	• Programming • Communication • Artificial intelligence • Data wrangling skills

What is Big Data ?

Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. But it's not the amount of data that's important. It's what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves.

While the term "big data" is relatively new, the act of gathering and storing large amounts of information for eventual analysis is ages old. The concept gained momentum in the early 2000s when industry analyst Doug Laney articulated the now-mainstream definition of big data as the three Vs

- ➔ **Volume.** Organizations collect data from a variety of sources, including business transactions, social media and information from sensor or machine-to-machine data. In the past, storing it would've been a problem – but new technologies (such as Hadoop) have eased the burden.
- ➔ **Velocity.** Data streams in at an unprecedented speed and must be dealt with in a timely manner. RFID tags, sensors and smart metering are driving the need to deal with torrents of data in near-real time.
- ➔ **Variety.** Data comes in all types of formats – from structured, numeric data in traditional databases to unstructured text documents, email, video, audio, stock ticker data and financial transactions.

Why is Big data Important?

The importance of big data doesn't revolve around how much data you have, but what you do with it. You can take data from any source and analyze it to find answers that enable 1) cost reductions, 2) time reductions, 3) new product development and optimized offerings, and 4) smart decision making. When you combine big data with high-powered analytics, you can accomplish business-related tasks such as:

- Determining root causes of failures, issues and defects in near-real time.
- Generating coupons at the point of sale based on the customer's buying habits.
- Recalculating entire risk portfolios in minutes.
- Detecting fraudulent behavior before it affects your organization.

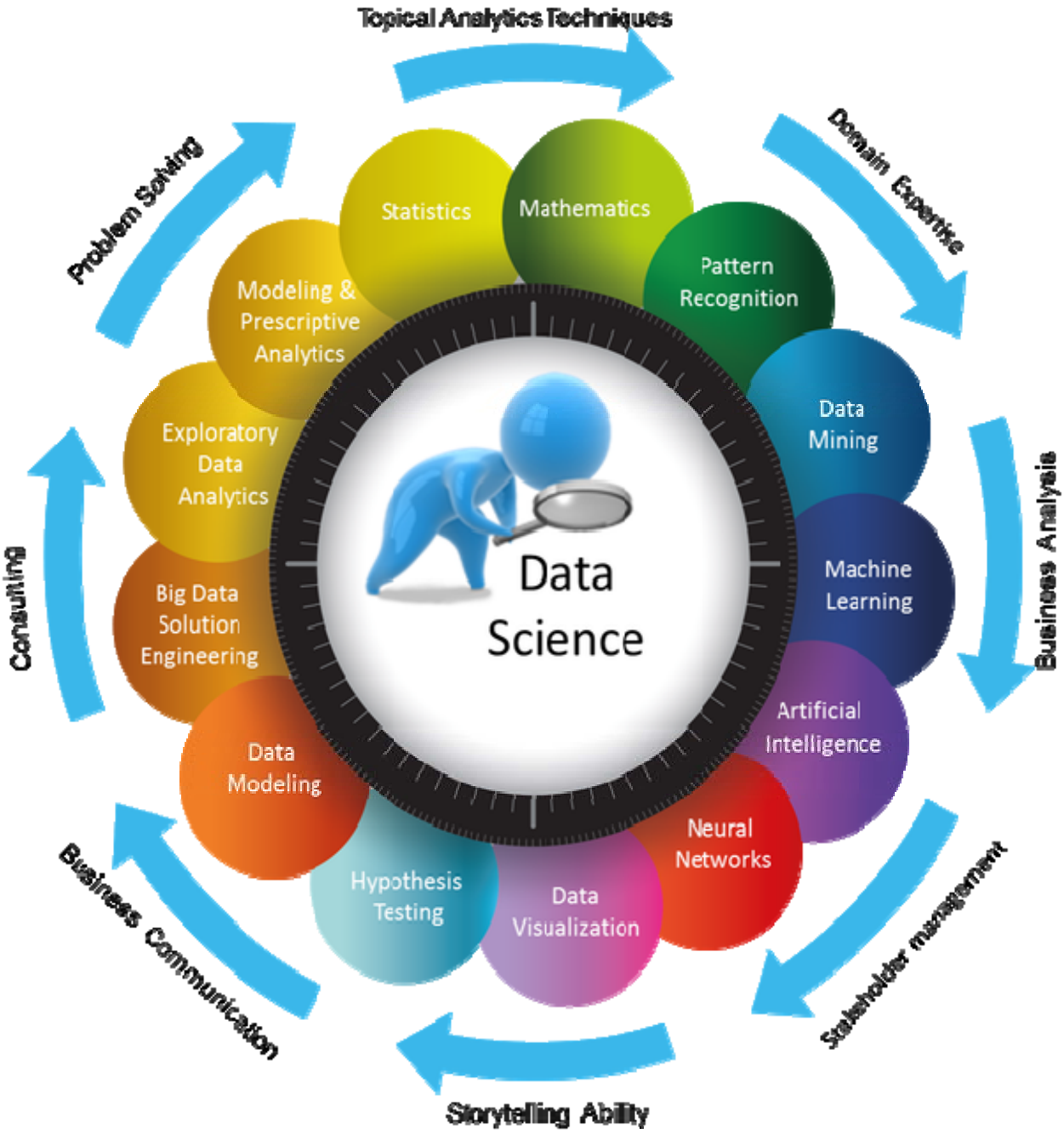
What is Traditional data ?

Traditional data systems, such as relational databases and data warehouses, have been the primary way businesses and organizations have stored and analyzed their data for the past 30 to 40 years. Although other data stores and technologies exist, the major percentage of business data can be found in these traditional systems. Traditional systems are designed from the ground up to work with data that has primarily been structured data.

Every year organizations need to store more and more detailed information for longer periods of time. Increased regulation in areas such as health and finance are significantly increasing storage volumes. Expensive shared storage systems often store this data because of the critical nature of the information. Shared storage arrays provide features such as striping (for performance) and mirroring (for availability). Managing the volume and cost of this data growth within these traditional systems is usually a stress point for IT organizations. Examples of data often stored in structured form include Enterprise Resource Planning (ERP), Customer Resource Management (CRM), financial, retail, and customer information.

Atomicity, Consistency, Isolation, Durability (ACID) compliant systems and the strategy around them are still important for running the business. A number of these systems were built over the years and support business decisions that run an organization today. Relational databases and data warehouses can store Peta bytes (PB) of information. However, these systems were not designed from the ground up to address a number of today's data challenges. The cost, required speed, and complexity of using these traditional systems to address these new data challenges would be extremely high.

Data Science Life cycle



What is Cybercrime ?

Cybercrime is a type of crime or an illegal activity that is basically committed through a computer with the help of networking.[1]

The ideal or most familiar word for this would be **"Hacking"**.

- Following are the examples of cybercrime.
  - Internet Fraud.
  - Cyberbullying.
  - Gathering Information Illegally.
  - Identity Theft.
  - Phishing scams.
  - Hate Crimes.

INTERNET FRAUD:

Whenever one tries to purchase a product from the internet, he/she is on a great risk of being victimized by internet fraud. So, it is very important for a person to research and study the sources from which he/she is purchasing the product. The e-commerce environment is stuffed with fake companies and organizations who distribute worst quality products to the customers and are greatly involved in stealing customers bank account data. These Internet crimes became a larger platform for crimes in the late 1990s and early 2000s.

CYBERBULLYING:

Cyberbullying is done through digital devices for example Cell Phones, Computers, Tablets, iPhone, I pads etc. Mentally torturing, harassment, Humiliation is cyberbullying. It includes sending, posting or sharing taunting, offending and sexual contents "publicly" or sending it to a specific person. The content could be personal data like pictures, email, chats etc.

GATHERING INFORMATION ILLEGALLY:

Physical **security** threats are often underestimated in favor of technical threats such as phishing and malware. Physical device threats occur when someone is able to physically gain access to your confidential data like data **gathered** from stolen devices.

NENTITY THEFT:

Identity theft, also known as **identity fraud**, is a crime in which an imposter obtains key pieces of personally identifiable information, such as Social **Security** or driver's license numbers, in order to impersonate someone else.

PHISHING SCAMS:

This is an attempt to gather users password and sensitive information by generating and disguising a fake webpage of an authentic website. When a user comes across this kind of fake webpage he inputs his personal data in the sign in or sign up form resulting in stolen data of the user.

HATE CRIMES:

A **hate crime** (also known as a bias-motivated **crime** or bias **crime**) is a prejudice-motivated **crime** which occurs when a perpetrator targets a victim because of his or her membership (or perceived membership) in a certain social group or race.

1.2 PROBLEM STATEMENT :-

- DEFINATION :

Cybercrime is defined as a crime in which a computer is the object of the crime (hacking, phishing, spamming) or is used as a tool to commit an offense (child pornography, hate crimes). Cybercriminals may use computer technology to access personal information, business trade secrets or use the internet for exploitative or malicious purposes. Criminals can also use computers for communication and document or data storage. Criminals who perform these illegal activities are often referred to as hackers.[1]

- PREDICTION :

The prediction of cyber crime data set are:

- From this prediction we will get to know how age wise crime is done.
- From this prediction we will also get to know what percent of crime happening in states of India.
- From this prediction we will get to know which age group commits more crime and in which states.
- From this prediction we will get to know in which year more crimes are committed and in which states.

- INPUT :

- In Cyber Crime dataset we have taken Year , Ages and State as input.

- OUTPUT :

- In Cyber Crime dataset we are defining how different age groups commits crime in different states of India.

- In Cyber Crime dataset we get yearly crimes committed by different age groups in different states in India.

➢ ADVANTAGES OF INDENTIFY THE DATASET :

- ### Interact with Data

A chief benefit of data visualization is that it brings exposes changes in a timely manner. But unlike static charts, interactive data visualizations encourage users to explore and even manipulate the data to uncover other factors. This creates a better attitude for use of analytics.

For instance, big data visualization tools can show a boat manufacturer that sales of its larger craft are down. This could be due to a number of reasons. But team members actively exploring related issues and correlating them to actual boat sales can identify the root causes and find ways to minimize their impact to drive more sales.

- ### Create New Discussion

One advantage to big data visualization is that it provides a ready means to tell stories from the data. Heat maps can show the development of product performance over time in multiple geographic areas, making it easier to see those that are performing very well or underperforming. This allows executives to drill down into specific locations to see what's being done well or poorly.

They may learn that targeting higher income market segments doesn't sell higher-priced products, or that traditionally solid sales of cleaning products are now less popular compared to environmentally-friendly green products. These insights could be used to brainstorm marketing strategies by region to support higher sales overall.

- ### Easy Comprehensive of Data

Utilizing data-visualization, companies may approach huge data and makes it easily comprehensible, be it the field of entertainment, current affairs, financial issues or political affairs. It also builds in them a deep insight, prompting them to take a good decision and an immediate business action if needed.