INTRODUCTION TO DATA SCIENCE

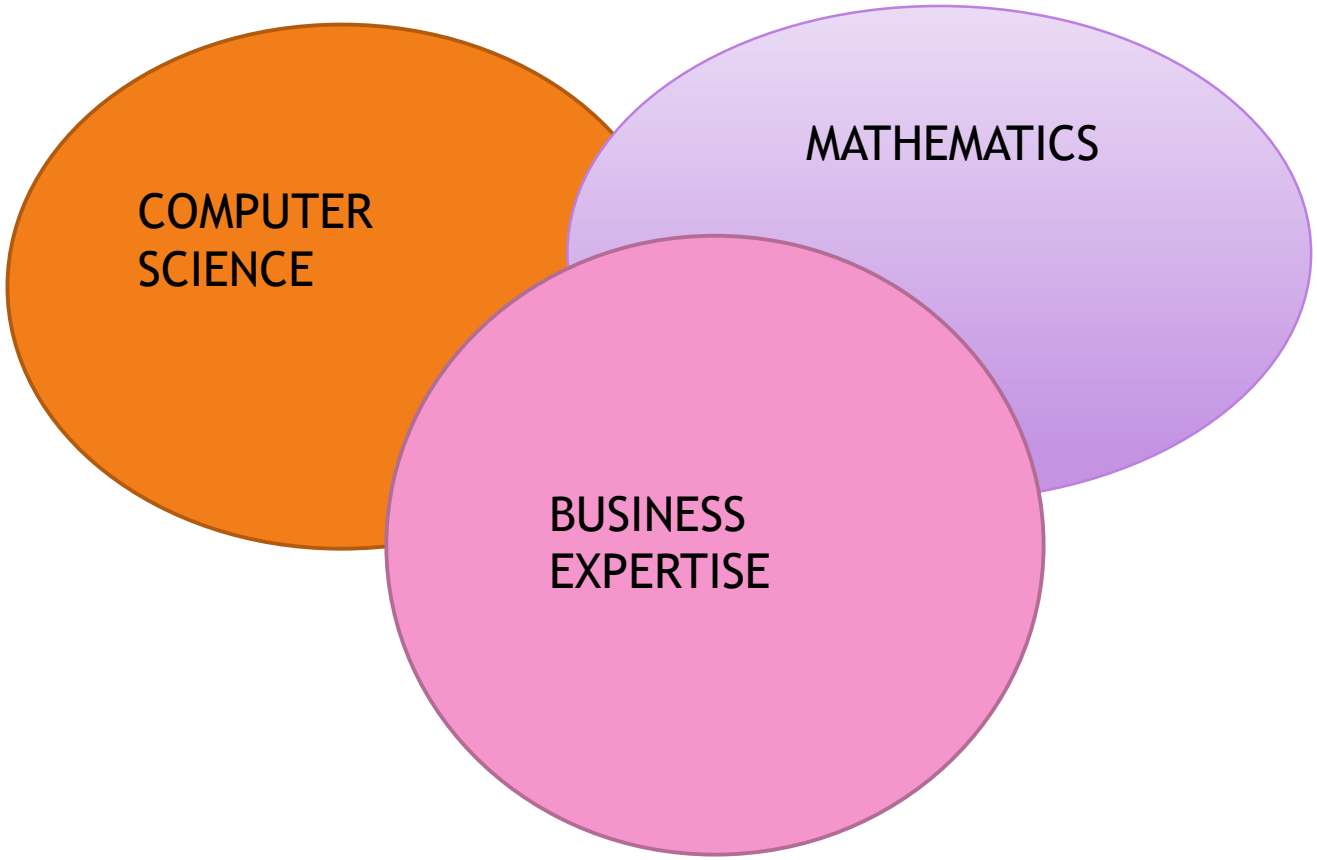

AI
DATA
SCIENCE

WHAT IS DATASCIENCE?

- Data Science is the field of study that involves extracting knowledge and insights from the noisy data and then turning those insights into actions that our business or an organization can take.

DIFFERENT AREAS COVERED BY DATA SCIENCE

- Data Science is the intersection among three disciplines.



Contt..

- True Data Science needs collaboration of all these three different areas.

TYPES OF DATASCIENCE

- DESCRIPTIVE ANALYTICS

- It involves having accurate data collection to make su what's happening in the business.

TYPES OF DATASCIENCE

- DIAGNOSTIC ANALYTICS
- This is more about why did something happened?
- Example, why did sales go up or down in a business?

TYPES OF DATASCIENCE

- PREDICTIVE ANALYTICS
- This is about what is likely to happen next?
- It involves using historical patterns in our data to predict the future outcomes.
- Example: What will be the sales in the next quarter?

TYPES OF DATASCIENCE

- PRESCRIPTIVE ANALYTICS
- What is the recommended best action for a particular outcome?
- Example, What needs to be done in order to improve sales by 20%?

DATA SCIENCE LIFE CYCLE

- BUSINESS UNDERSTANDING
- Having the business and domain expertise can make sure that we are asking the right question.
- DATA MINING
- Data mining is the process of sorting through large data sets to identify patterns and relationships that can help solve business problems through data analysis.

DATA SCIENCE LIFE CYCLE

- DATA CLEANING

- When we find data its probably not in the best format i.e it may contain duplicates , missing values , so data cleaning is required before analysis.

- EXPLORATION
- It will help to answer the different questions that were asked earlier.

DATA SCIENCE LIFE CYCLE

- We use advance machine learning tools to make preductions and prescribe actions in the future.
- VISUALIZATION
- Now there is need to visualize our insights and outcomes for analysis.

Where is Data Science Needed?

- For route planning: To discover the best routes to ship
- To foresee delays for flight/ship/train etc. (through predictive analysis)
- To create promotional offers
- To find the best suited time to deliver goods
- To forecast the next years revenue for a company
- To analyze health benefit of training
- To predict who will win elections

How Does a Data Scientist Work?

- A Data Scientist requires expertise in several backgrounds:
- Machine Learning
- Statistics
- Programming (Python or R)
- Mathematics
- Databases

Here is how a Data Scientist works:

- **Ask the right questions** - To understand the business problem.
- **Explore and collect data** - From database, web logs, customer feedback, etc.
- **Extract the data** - Transform the data to a standardized format.
- **Clean the data** - Remove erroneous values from the data.

Contt..

- **Find and replace missing values** - Check for missing values and replace them with a suitable value (e.g. an average value).
- **Normalize data** - Scale the values in a practical range (e.g. 140 cm is smaller than 1,8 m. However, the number 140 is larger than 1,8. - so scaling is important).
- **Analyze data, find patterns and make future predictions**.
- **Represent the result** - Present the result with useful insights in a way the "company" can understand.

TYPES OF DATA

- Structured data
- Unstructured Data and Semistructured Data.

STRUCTURED DATA

- Structured data adheres to a pre-defined data model. This model describes how data is recorded, and it defines the attributes and provides information about the data type (e.g. name, date, number) and restrictions on their values (e.g. number of characters).

- Example, phone numbers,heights or weights , product names etc.

UNSTRUCTURED AND SEMISTRUCTURED DATA

- Unlike structured data, unstructured data requires human interpretation. Consider a block of text. Computers can read each word, or sentence, but they can’t (yet) determine the meaning or tone of the text without human intervention.

- Example, images (human- and machine-generated)
- video files
- audio files
- social-media posts
- product reviews

UNSTRUCTURED AND SEMISTRUCTURED DATA

- Some data, such as email, is considered to be semi-structured. Email headers contain metadata such as the date, language, and recipient's email address, which are all structured data. But the email body, which contains your message, is unstructured.

Major Applications of Data Science

- In Search Engines
- In Transport
- In Finance
- In E-Commerce
- In Health Care
- Image Recognition
- Targeting Recommendation
- Autocomplete

IN SEARCH ENGINES

- The most useful application of Data Science is Search Engines. As we know when we want to search for something on the internet, we mostly used Search engines like Google, Yahoo, Safari, Firefox, etc. So Data Science is used to get Searches faster.

In Transport

- In Driverless Cars the training data is fed into the algorithm and with the help of Data Science techniques, the Data is analyzed like what is the speed limit in Highway, Busy Streets, Narrow Roads, etc. And how to handle different situations while driving etc.

In Finance

- Data Science plays a key role in Financial Industries. Financial Industries always have an issue of fraud and risk of losses. Thus, Financial Industries needs to automate risk of loss analysis in order to carry out strategic decisions for the company.

In Ecommerce

- E-Commerce Websites like Amazon, Flipkart, etc. uses data Science to make a better user experience with personalized recommendations.

In HealthCare

- In the Healthcare Industry data science act as a boon. Data Science is used for:
- Detecting Tumor.
- Drug discoveries.
- Medical Image Analysis.

Image Recognition

- Currently, Data Science is also used in Image Recognition. **For Example,** When we upload our image with our friend on Facebook, Facebook gives suggestions Tagging who is in the picture.

Targeting Recommendation

- Targeting Recommendation is the most important application of Data Science. Whatever the user searches on the Internet, he/she will see numerous posts everywhere.

Auto Complete

- AutoComplete feature is an important part of Data Science where the user will get the facility to just type a few letters or words, and he will get the feature of auto-completing the line.

THANKS