**Lecture- Data Analytics**

**What is Data Analytics:**

* + Process of Analysing Raw Data to Get insights and find Patterns About the Information
  + Using several Techniques and Processes The raw data is transformed into a Format for further process, Visualized and be made Human Readable.
  + Data Analysis is nothing but finding insights from Data to make human life easy
  + 90’s Retail store and todays D-mart
* We have Data All over
* Structured data unstructured data
* Data Analyst, analyses data to create structure/Order/Meaning out of it.

1. Descriptive- Creating Dashboard for Daily Performance of Product.

But they Do not tell us why change happens or what will happen in Future

1. Predictive- Utilizing Historical Data, predict which will making more Profit.

Which Customer should we show advertisement

Which movie will be released on social media platform

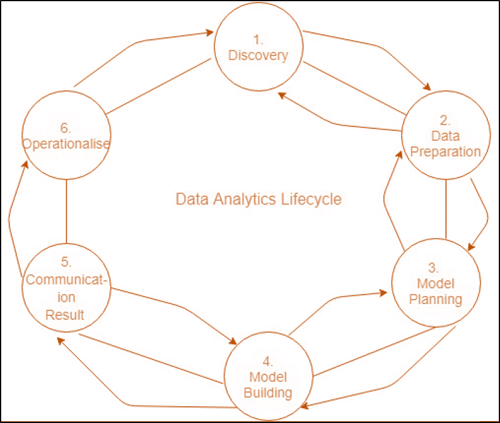
1. Prescriptive- How Much Discount offer on Which platform or which platform to use to engage a customer and which contains to showcase

Key Differences Between Data Science and Artificial Intelligence

* Purpose: Data science is mainly focused on extracting insights and information from vast amounts of data. AI, on the other hand, focuses on creating systems that can perform tasks without any explicit instructions.
* Scope: Data science encompasses various techniques from statistics, data analysis, and ML to analyse and interpret complex data. AI is broader and includes areas like robotics, natural language processing, and more.
* Tools: Data scientists often use tools like Python, R, SQL, and platforms like Jupyter and Tableau. AI researchers and developers might use TensorFlow, PyTorch, or OpenAI's platforms.
* Implementation: While data science often concludes with insights and decisions, AI aims for automation and creating systems that can act on their own.

## ****Data Analytics Life Cycle?****

The circular process of the Data Analytics Life Cycle consists of six key steps that govern how information is created, collected, processed, used, and evaluated. Setting company goals and working toward them will guide you through the remaining stages.



The Data analytics lifecycle was designed to address Big Data problems and data science projects. The process is repeated to show the real projects. To address the specific demands for conducting analysis on Big Data, the step-by-step methodology is required to plan the various tasks associated with the acquisition, processing, analysis, and recycling of data.

### **Phase 1: Discovery -**

* The data science team is trained and researches the issue.
* Create context and gain understanding.
* Learn about the data sources that are needed and accessible to the project.
* The team comes up with an initial hypothesis, which can be later confirmed with evidence.

### **Phase 2: Data Preparation -**

* Methods to investigate the possibilities of pre-processing, analysing, and preparing data before analysis and modelling.
* It is required to have an analytic sandbox. The team performs, loads, and transforms to bring information to the data sandbox.
* Data preparation tasks can be repeated and not in a predetermined sequence.
* Some of the tools used commonly for this process include - Hadoop, Alpine Miner, Open Refine, etc.

### **Phase 3: Model Planning -**

* The team studies data to discover the connections between variables. Later, it selects the most significant variables as well as the most effective models.
* In this phase, the data science teams create data sets that can be used for training for testing, production, and training goals.
* The team builds and implements models based on the work completed in the modelling planning phase.
* Some of the tools used commonly for this stage are MATLAB and STASTICA.

### **Phase 4: Model Building -**

* The team creates datasets for training, testing as well as production use.
* The team is also evaluating whether its current tools are sufficient to run the models or if they require an even more robust environment to run models.
* Tools that are free or open-source or free tools Rand PL/R, Octave, WEKA.
* Commercial tools - MATLAB, STASTICA.

### **Phase 5: Communication Results -**

* Following the execution of the model, team members will need to evaluate the outcomes of the model to establish criteria for the success or failure of the model.
* The team is considering how best to present findings and outcomes to the various members of the team and other stakeholders while taking into consideration cautionary tales and assumptions.
* The team should determine the most important findings, quantify their value to the business and create a narrative to present findings and summarize them to all stakeholders.

### **Phase 6: Operationalize -**

* The team distributes the benefits of the project to a wider audience. It sets up a pilot project that will deploy the work in a controlled manner prior to expanding the project to the entire enterprise of users.
* This technique allows the team to gain insight into the performance and constraints related to the model within a production setting at a small scale and then make necessary adjustments before full deployment.
* The team produces the last reports, presentations, and codes.
* Open source or free tools such as WEKA, SQL, MADlib, and Octave.