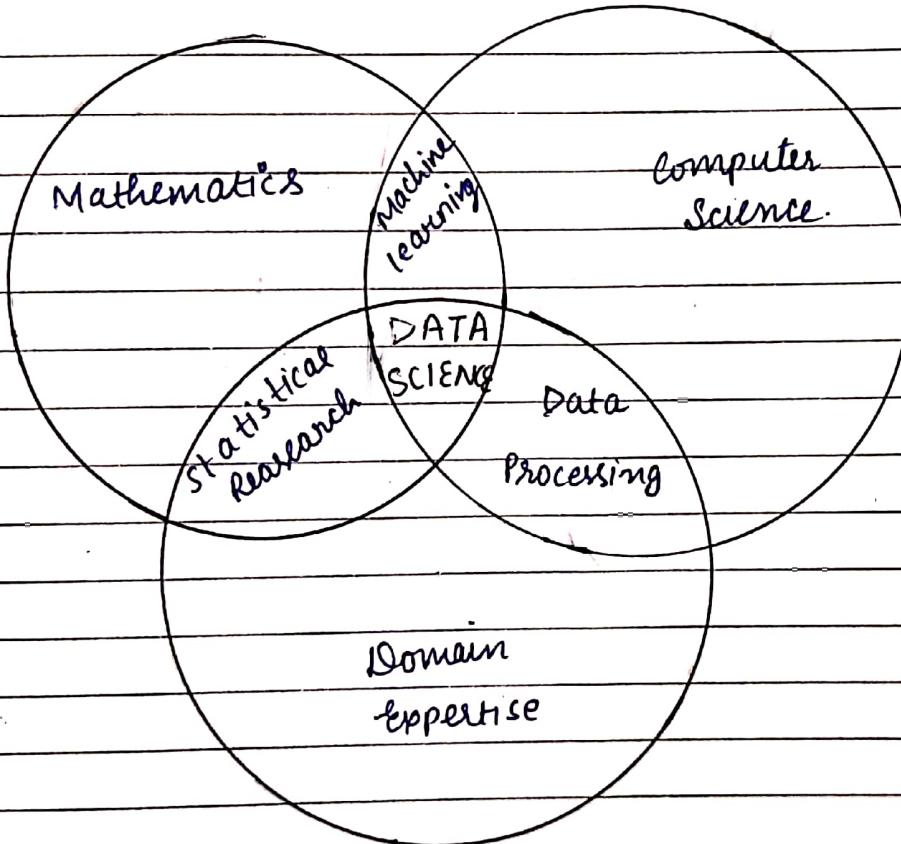


TOPIC ..... DATE .....

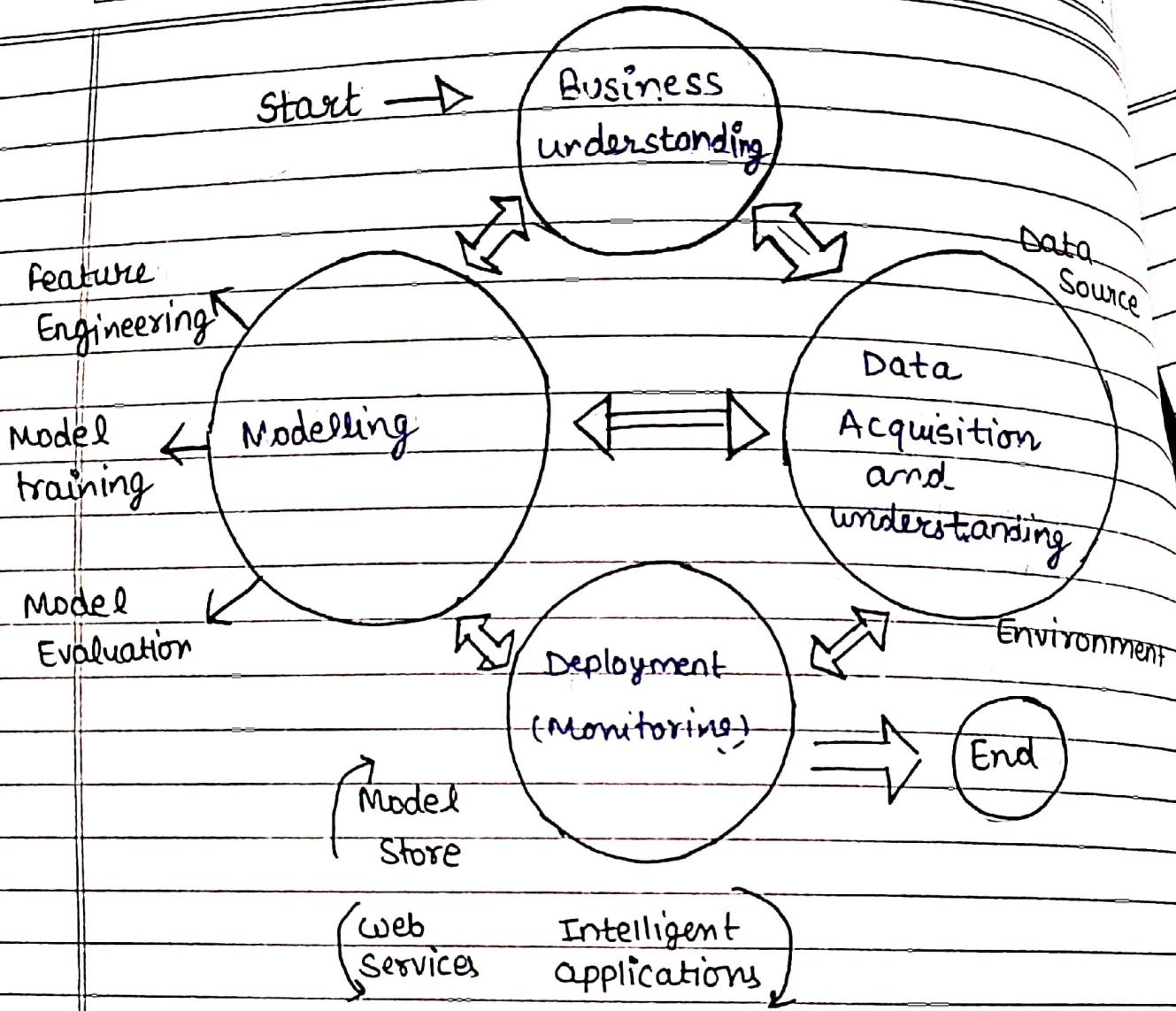
## Assignment - I

Ques 1 Explain Life cycle of Data science and Application Domains of Data science ?

Ans Data science is a multidisciplinary field that utilizes techniques from computer science, statistics, and domain knowledge to extracts insights, patterns from knowledge from data.



A data science lifecycle indicates the iterative steps taken to build, deliver and maintain any data science product. A general data science lifecycle process includes the use of machine learning algorithms and statistical practices that result in better predictions model.



- Business understanding - To define the Business understanding for data science project following are essential points
  - # List the issues that needs to be resolved.
  - # Define the Project's potential value.
  - # Determine the project's risks taking ethical issues into account.
  - # create and distribute a flexible , high level project plan.
- Data Acquisition and understanding -
  - # Selecting data Related to the problem.
  - # combining the data sets , you may integrate the data .
  - # clean data to find missing values.

# Handle the missing values by removing or inputting them.  
# Errors are dealt with by being removed.  
# use the box plots for detecting outliers & handling them.

- Modelling - After the data has been analyzed and visualized the next critical step is data modelling. The key components are kept in dataset and therefore the data gets refined. Because there are numerous method for modelling data, it is critical to determine which one is most effective. The model is now being tested in Real world data. where there are few data points, the output monitored for improvement. while the model is being evaluated or tested, data may change, and output may alter dramatically.
- Deployment : In this step, the delivery method that will be used to distribute the model to users or another system is created. If any step is performed improperly, and hence, have an effect on subsequent step and complete effort goes to waste.

It concludes that the Data science lifecycle, focusing on applying machine learning and Analytical methodologies to achieve business goals.

### Applications Domains of Data Science -

# Data Science

- 1. Fraud Detection
  - 2. credit Risk modelling
  - 3. customer lifetime value.
- 
- 1. Identifying consumers
  - 2. Recommending products
- 
- 1. Anomaly Detection
  - 2. Maintenance Detection.
  - 3. Monitoring systems
  - 4. Predicting Potential Problems.

Banking

E-commerce

Manufacturing

Health care

Transport

Finance

Medical image analysis  
Drug Discovery

Bioinformatics  
virtual assistance  
self Driving cars  
car monitoring system

Enhanced Driving  
Enhanced safety

customer segmentation  
strategic decision making

Algorithmic trading  
Risk analysis.

with each passing day, the volume of data is increasing and it demands analytical tools for storing, processing and analyzing data.

**Question:** what are the skills needed to become a data scientist and also explain roles and responsibilities of a data scientist?

**Answer:** There are two types of skills needed to become a data scientist -

TOPIC ..... DATE .....

### TECHNICAL SKILLS

- \* Machine learning
- \* Deep learning
- \* Programming Language
- \* Statistics.
- \* Big data.
- \* Mathematics.
- \* Data visualization.
- \* Knowledge of MS Excel

### NON TECHNICAL SKILLS

- \* Communication skills.
- \* Business Acumen.
- \* Great data intuition.

### Roles and Responsibilities of Data Scientist -

A Data scientist is a professional who collects large amount of data using analytical, statistical and programmable skills.

- \* Analyzing large amount of information to find patterns and solutions.
- \* Collaborate with Business and IT teams.
- \* Presenting results in clear manner.
- \* Developing prediction system and machine learning algorithms.
- \* Carrying out preprocessing of structured and unstructured data.
- \* Data mining or Extracting usable data from valuable data sources.
- \* Collect data and identify data sources.
- \* Create solutions and strategies to business problem.
- \* Present data using various techniques and tools.

Ques3

what is the importance of python in Data Science ?

Answer3

There are numerous reasons why data python is important in data science -

- \* Flexibility - python is a simple language and not require as much efforts in term of coding as other programming language.
- \* well-supported - There are many analytics libraries available in python.
- \* Easy to learn - Python has a simple and intuitive syntax that is easy to learn and Read.
- \* Versatility - It allows data scientists to perform a wide range of tasks.
- \* Data handling - Python excels at data manipulation and cleaning, which are critical steps in data science workflow. Libraries like pandas provide powerful tools for data wrangling, including data loading, filtering, aggregation and transformation.
- \* Reproducibility - It supports reproducibility in projects through tools like Jupyter notebook, which allow researchers to document and share work with interactive code and Explanations.
- \* Integration : Python can be seamlessly integrated with other data science tools and technologies for example - It can connect to databases, web APIs and Big data processing framework like Hadoop and spark.

Ques 4 Structured vs Unstructured Data: Compare & Explain.

Answer Structured data stands for information that is highly organized, factual, and to the point. It usually comes in the form of letters and numbers that fit nicely into the rows and columns of tables. Structured data commonly exists in tables similar to Excel files and Google Docs Spreadsheet.

Unstructured data doesn't have any predefined structure to it and comes in all its diversity of forms. The examples of unstructured data vary from imagery and text files like PDF documents to video and audio files to name a few.

S.NO	STRUCTURED DATA	UNSTRUCTURED DATA
1.	Data that is collected from a known method and can be neatly arranged.	Data has no pre-defined format or organization, making analysis difficult.
2.	Graphical Representation -	Graphical Representation
		
3.	Storage location - stored in warehouse	Storage location stored in data lakes.

TOPIC ..... DATE .....

4. Storage size -

- Requires less storage

Storage size -

- Requires a lot of storage.

5. uses -

machine learning

uses -

Natural language processing

6. Data nature -

Quantitative

Data nature -

Qualitative.

7. formats -

Several formats

formats -

huge variety of formats