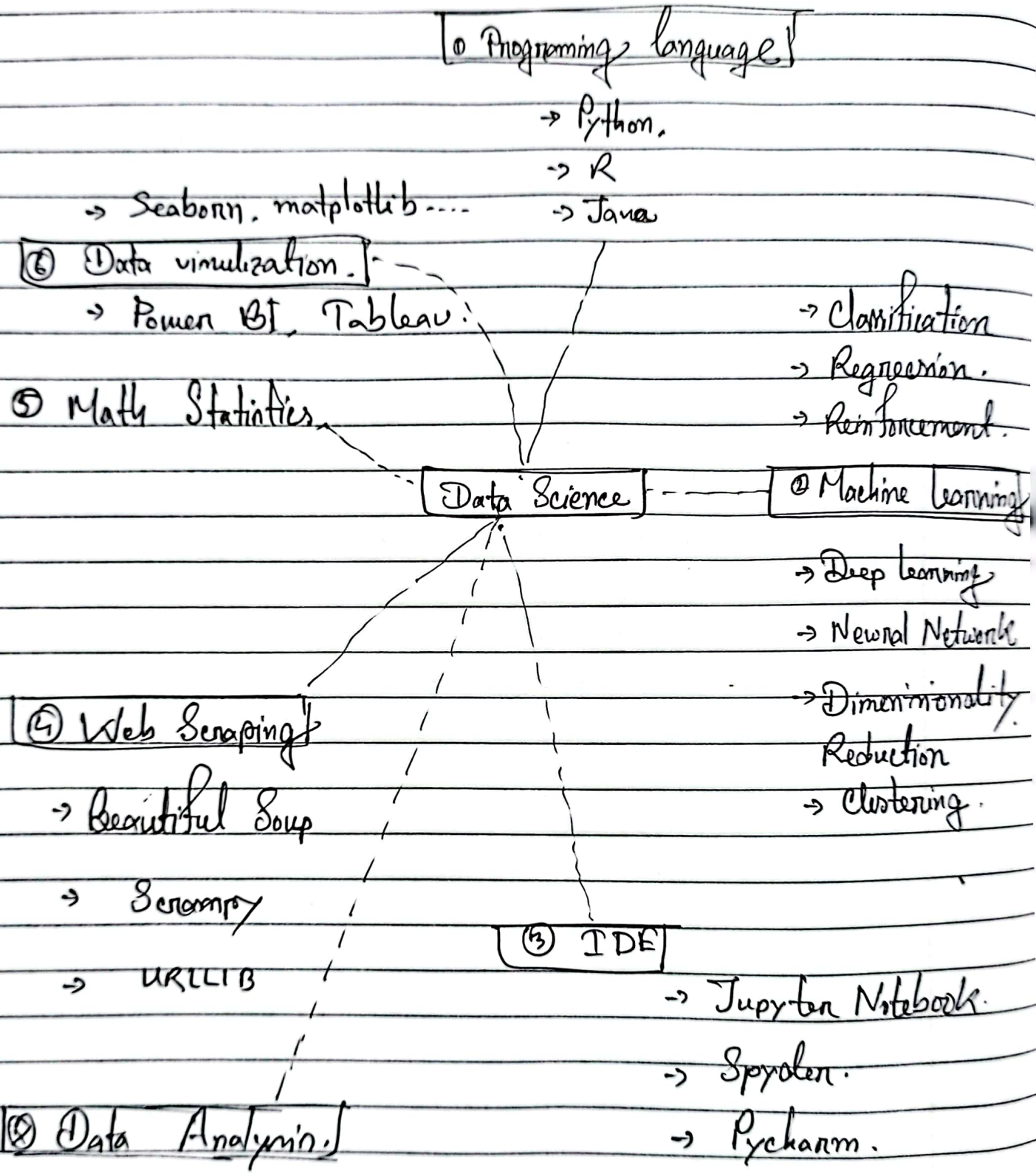

① Programming language
→ Python.
→ R
→ Java
→ Seaborn, matplotlib....
⑥ Data visualization.
→ Power BI, Tableau.
⑤ Math Statistics.
Data Science
② Machine learning
→ Classification
→ Regression.
→ Reinforcement.
→ Deep learning
→ Neural Network
→ Dimensionality Reduction
→ Clustering.
④ Web Scraping
→ Beautiful Soup
→ Scrapy
→ URLLIB
③ IDE
→ Jupyter Notebook.
→ Spyder.
→ Pycharm.
⑦ Data Analysis.
→ Feature Engineering
→ Data Wrangling
→ EDA

Classification — fixed category

→ Binary classification. (2 cat)

→ Multiclass classification. (more than 2)

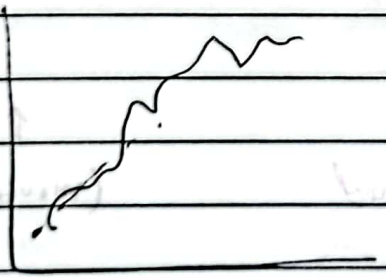
→ Multilabel classification

↓

movie

	Action	Romance	Thriller	Com
Endgame	1	0	1	1
Drishyam	0	0	1	1

forecasting → Regression.



time Series analysis.

Regression → Predict Continous value.

Unsupervised Machine learning.

→ Clustering

→ Segmentation

→ Reduce Dimension

Semi-Supervised ML

label data → Model → Recommendation (o/p)

→ Reinceforment learning

continously learining and new Recommendation.

Ex:- Children learning way is semi-supervised or Reinceforment learning.

(movie recommendation)

Step ↓

① Programming Language

→ Python

→ R

→ Julia.

② Exploratory Data Analysis.

③ Feature Engineering.

④ Feature Selection.

⑤ Machine learning Algorithm.

⑥ Regression, Classification, clustering.

Linear, Logistic Regression, Decision tree, Random Forest, K-means ---------

⑦ Hyperparameter Tuning →

GridSearch, RandomizedSearch, Hyperopt, Optuna, Genetics Algorithms.

⑨ Docker And Kubernetes.

⑩ Model Deployment.

⑪ End to End ML Projects.

Google Cloud

Google colab

AWS

Azure.

Types of Encoding

→ categorical variable

as like. Gender.

female
Male

Categorical variable

① Nominal Encoding

→ Don't worries about order.

technique

① → One hot encoding.

② → One hot encoding many categorical.

→ select top freq feature than apply one hot enc.

③ → Mean Encoding.

→ If. categories = 50
top freq. feature = 10
than column = 9

② Original Encoding

→ Rank or Order matter.

BE	PHD-1
BCOM	MAS-2
PHD	BAL-3
MASTER	BCOM-4

① → label Encoding.

② → Target guided Ordinal Encoding.

One hot encoding

→ Dummy variable

→ 4 Category

4-1 = 3 column

→ 5 Category.

5-1 = 4 Column

One column skiped

Suppose State
German
Spain
~~france~~

German	Spain	france
1	0	0
0	0	1
0	1	0

we can skip one column.

④ Why we need feature scaling?

When and why

we should Perform feature Scaling.

And when should not — why not

⇒ Handle missing values of categorical variable:

① Ignore the observation or Delete the row.

if we have large dataset → if not this is not good way.

② Build a model to predict the missing value
(can apply classifier algorithm to predict.)

③ Replace with the most frequent values.

④ Apply unsupervised learning.
using an algorithm like KNN.

High cardinality -> In a Column -> Categorical variable

- high num of unique values.
- high label category.

Related to the time series data.

Decision Tree.

classification

$$Entropy(S) = \sum_{i=1}^{n} - P_i \log_2 P_i$$

$$Gain(S, A) = Entropy(S) - \sum Entropy(S, A)$$

(Σ Prob × Entropy)

focus → higher Information Gain = more entropy remove

Partition - more আলাদা দেখাবে → Target variable.

Partition " বেশি দেখাবে →

than আগেরটা

Information Gain এর বেশি → সেটা root node.