***5th September 2023***

* Data V/S Information:
  + Example: sound of train
  + Personal example: right at this moment, it is 12:00, it does not matter as it is just a representation of fact, my next class is at 6.30, now this is the information.

***19th September 2023:***

Human has two modes of thinking:

1. Intuition (System 1) – we have some experience dealing with the situation.
2. Thinking (System 2) - we plan, analyze and then do something.

Why should we keep people in the loop?

* Commonsense: so that they can have a buy in
* Empathy: to make sense of how it would feel after a decision is taken
* Abstraction: summarizing things without adding in more details.

**Cross Industry Standard Process (CRISP-DM)**

**What Tasks Data Mining Can Accomplish?**

6 phases: (First three are not part of the course)

1. Description

Describe patterns or trends in data (drink and drive example, California police patrol example)

1. Estimation

Estimate some information from the given data (average annual income of salaried class example)

1. Prediction

as

1. Classification

*Claritas website*07450

1. Clustering
2. Association

Data validity should not be associated with time.

***26th September 2023: (KNN)***

* Scaling creates a huge impact on distance.
* This course is concerned with Min-Max Normalization using formula:

***17th October 2023 (Decision Tree):***

* It is better w.r.t KNN because data splitting is more concrete based on statistical evidence.
* The maximum value of Sigma function in case of binary split is going to be 2 e.g., all bad boys in one bucket and good boys in other bucket.
* Maximum value for Pl \* Pr is going to be 0.25 (Pl=.5 and Pr=.5)