Everything on What is Data Mining about?

Data mining is about predicting with no previous knowledge just data. Data is usually a snapshot of the world. Collecting data can have many problems:

1) **Errors**   : Data measured or entered wrong.
2) **Quantity**   : Data might not be enough to represent the problem
3) **Quality**   : Data might not contain all necessary information
4) **Cost**   : Some data is particularly expensive to collect or even unfeasible

Advantages

Data mining is capable of processing vast quantities of data. It also allows computers to learn how to perform tasks that we cannot program them to do.

Variables/Attributes

Variables can be : Numeric or Nominal (Categorical)

Numeric variables can be : **Discrete**   → Only certain values e.g Decimals not acceptable
                           Continuous → Any number is acceptable

Definitions

Data Model : A representation of some aspect of data. Data mining models train to generalise from specific data to general relationships

Learning : The process of building a model, using data, capable of performing a given task

Inference : The process of providing a new input pattern and the software producing an output pattern, The output is a prediction of classification