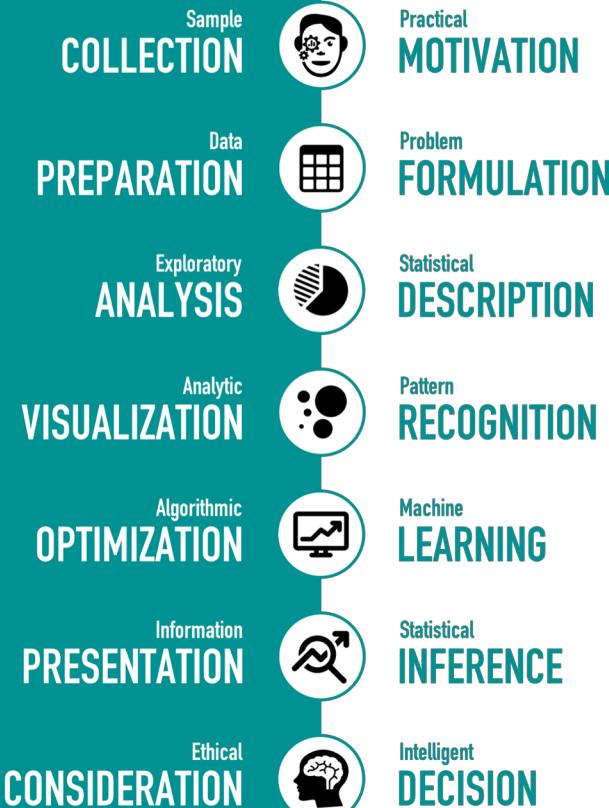


# Data Science Solutions

**Sourav SEN GUPTA**  
Lecturer, SCSE, NTU





## Data Science Common Problems

### Five Primary Questions

- How much? How many?
- Is it type A or type B?
- How is this organized?
- Is it a weird behavior?
- What should be done next?

<https://www.youtube.com/watch?v=0XyV91VYrDs>



# Data Science Common Problems

**Prediction : Numeric**

**How much?  
How many?**

What is the expected Sales of the  
next game of this game franchise?  
Is it profitable to make the sequel?



# Data Science Common Solutions

## Prediction : Numeric Regression

Try to find the relationship of Sales of the games with other Variables, like Graphics Quality, Genre, etc.

Model :  $\text{Sales} = f(\text{Variables})$



# Data Science Common Solutions

## Prediction : Numeric Regression

Model : Sales =  $f(\text{Variables})$

Linear Regression Models  
Tree Models for Regression  
Neural Network for Regression



# Data Science Common Problems

**Prediction : Classes**

**Is it type A  
or type B?**

What is the chance that a student  
will get into NTU in AY2019-2020?  
Will an application be successful?



# Data Science Common Solutions

**Prediction : Classes**

## Classification

Try to find the Probability of getting admitted to NTU in terms of other Variables, like Scores, Gender, etc.

Model :  $\mathcal{P}(\text{Admit}) = f(\text{Variables})$



# Data Science Common Solutions

## Prediction : Classes Classification

Model :  $\mathcal{P}(\text{Admit}) = f(\text{Variables})$

Logistic Regression Model  
Tree Models for Classification  
Neural Network for Classification



## Data Science Common Problems

**Detection : Structure**

# How is this organized?

Is there any structure apparent  
within the FairPrice customers?  
Which customer group to target?



## Data Science Common Solutions

### Detection : Structure Clustering

Try to find Groups of Data Points  
that are close together but are far  
from the other Groups of Points.

Close–Far depends on “Distance”



# Data Science Common Solutions

## Detection : Structure Clustering

Close–Far depends on “Distance”

Distance: Euclidean, Jaccard etc.  
k-Means Algorithm for Clustering  
Hierarchical Model for Clustering



# Data Science Common Problems

Detection : Anomaly

Is it weird  
behavior?

Is this Boeing engine behaving in unusual fashion during the flight?  
Is the engine still safe to operate?



# Data Science Common Solutions

**Detection : Anomaly**

## Anomaly Detection

Try to find Deviations of the Data compared to the Regular Pattern observed through the data model.

Deviations depend on the Model



# Data Science Common Solutions

Detection : Anomaly

## Anomaly Detection

Deviations depend on the Model

Cluster-Analysis based Detection  
Nearest Neighbor Detection Model  
Support Vector based Detection



# Data Science Common Problems

**Decision : Action**

## What should be done next?

Should brake at the Yellow Light or  
should the car accelerate instead?  
Which action will be rewarded?



# Data Science Common Solutions

**Decision : Action**

## Adaptive Learning

Try to model a Profit/Loss Function depending at any given state, and try to Maximize/Minimize the same.

Optimize  $f( \text{State, Variables} )$



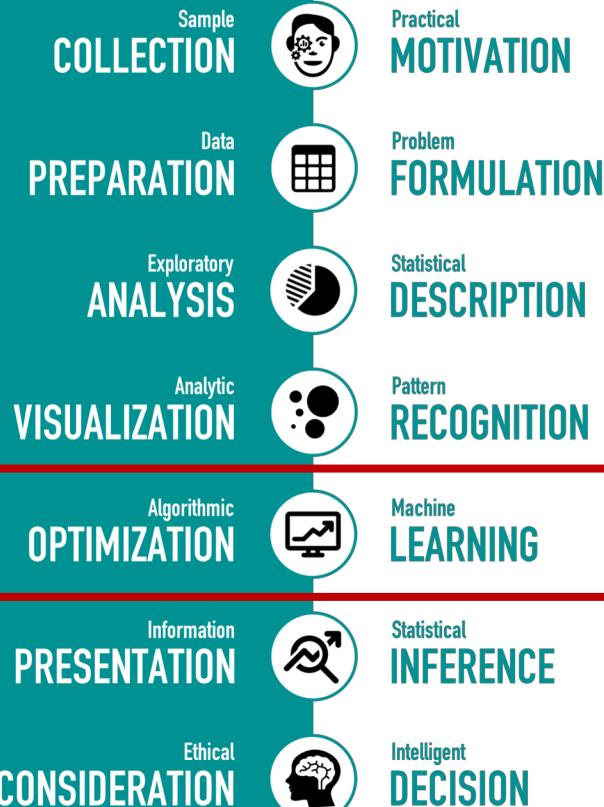
# Data Science Common Solutions

**Decision : Action**

## Adaptive Learning

Optimize       $f(\text{State, Variables})$

Reinforcement Learning Approach  
Monte-Carlo, State-Action-Reward  
Q-Learning, Deep Reinforcement



## Data Science Pipeline Algorithmic Solution

How to prepare the acquired Data?  
How to create a model for the Data?  
How to choose the optimal Model?

## How to intelligently infer from the Model?