Prof. Somnath Mandal - Data Science, AI
somenathmandal14@gmail.com

What is Data Science ?

'Full Stack Data Science' or 'End to End Data Science'

Big data → Hadoop (Engineering side) How to store & use huge amount of data

✓ Data Engineering → sql (write) queries → collect data, merge & give it to us

✓ Machine Learning → What do you with that data → model

Deployment → interface i.e creating and application
i.e : Create a application for customer to use
→ Flask, AWS, Heroku, Microsoft Azure.

Lifecycle of Data Science

Starts with a problem that adds value to the business

No Data Science problem

↓

only Business problem that is solved by Data Science.

ex : google earns from advertisment i.e Customer Specific
→ 'click on the advertisment'

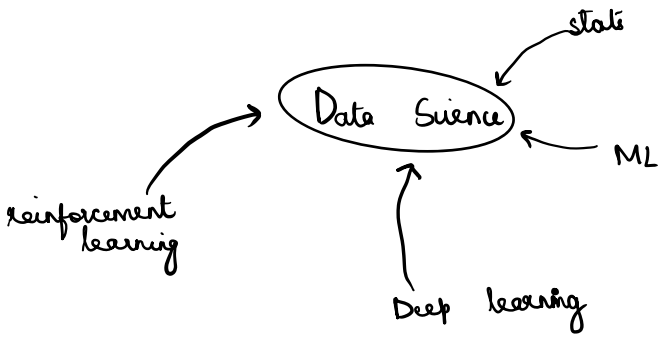
therefore they need to know you i.e your profile and preference.

per click they get 300 rupees.

Convert Business problem to Data Science problem

Data Science is a small problem of A.I

↳ Automation / reduce human effort using Data



step 1 : Think of possible causes

Step 2 : Check our data

is our assumption correct ?

step 3 : Where to get data ?

data Storage / hive

data Engineer → SQL query to get the data

Problems Data Quality

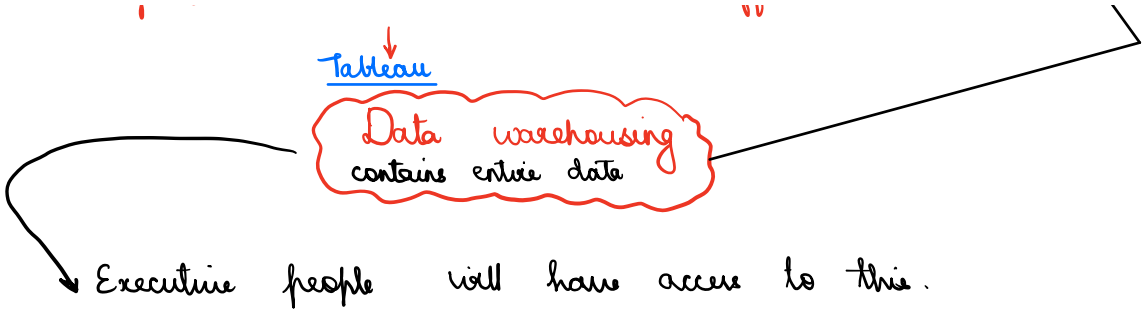
Data Mismatch

Record Missing

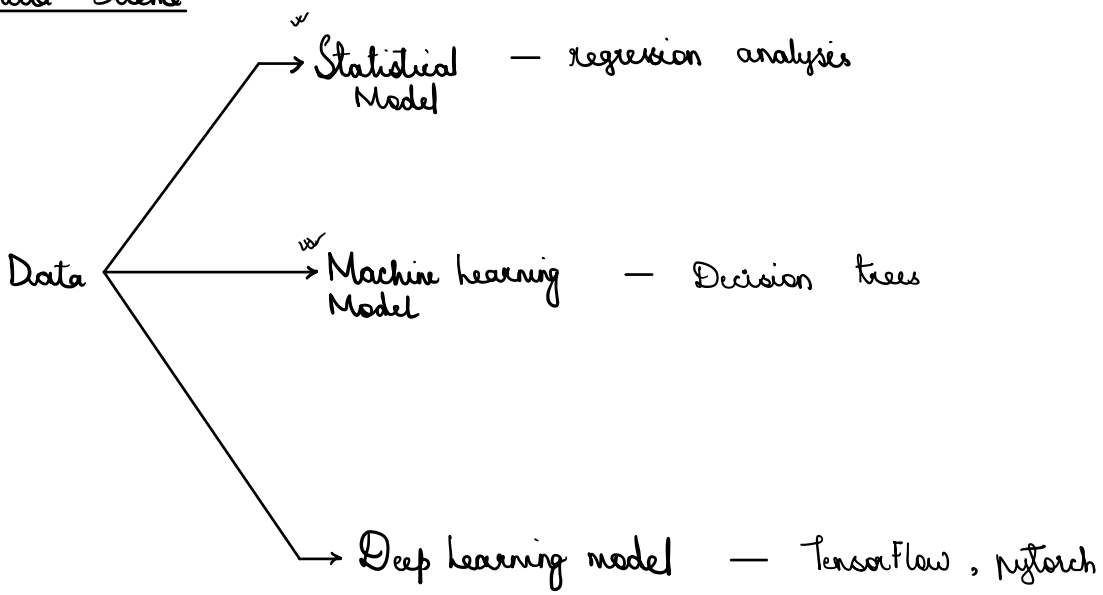
Extract Transfer Load → ETL

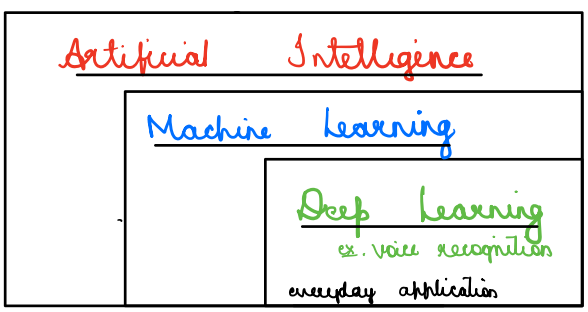
Database

Organisations collect data on different level



Data Science





imp

Deployment : Heroku, AWS, Flask, Microsoft Azure

↳ This is what the customer sees.

How do you come up with the reasons

Subject Matter Expert (SME)

↳ will give you reasons so you can collect data to corroborate with the assumption

I Probability

ex. Ludo

6 + 1
5 + 2
4 + 3
3 + 4
5 + 2
6 + 1
6

$6/36 = 1/6$

$\Rightarrow \frac{\text{No of cases}}{\text{total no of cases}} = \frac{6}{6 \times 6} = \frac{6}{36} \Rightarrow \text{Prob} = 1/6$

probability should not be greater than 1

Probability : $\frac{\text{Possible cases}}{\text{total no of cases}}$

$0 \leq P \leq 1$

ex 2. 2 coins → HT, TH } HH, HT, TH, TT = 2/4 = 1/2.

ex 3. 2, 3, 5 ⟶ 3/6 = 1/2 P(A) = 1/2

events

Mutually Exclusive events (MEE)

→ When something happens the other absolutely cannot happen.

ex 1: Toss a coin

A = { outcome is H } ∴ A & B are mutually exclusive events

B = { outcome is T }

ex 2: toss a dice

A = { outcome is an even number }

B = { outcome is an odd number }

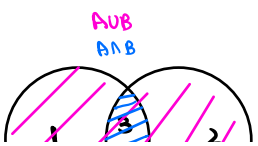
∴ A & B are mutually exclusive events

Probability of mutually exclusive events

set : It's a collection of well defined and unique objects.

ex: A = { 1, 3, 5 }

B = { 2, 3, 5 }



$A \cup B = \{1, 2, 3, 5\}$ ⇒ Union

$A \cap B = \{3, 5\}$ ⇒ Intersection

$A^c = U - A$, everything other than A

$B^c = U - B$, everything other than B

for mutually exclusive events

$P(A \cup B) = P(A) + P(B)$

$A = \{2, 4, 6\}$

$B = \{1, 3, 5\}$

$A \cup B = \{1, 2, 3, 4, 5, 6\}$ =

$P(A) = 3/6 = 1/2$

$P(B) = 3/6 = 1/2$

$P(A \cup B) = 6/6 = 1$

$P(A) + P(B) = 1/2 + 1/2 = 2/2 = 1 = P(A \cup B)$

Not mutually exclusive

$A = \{2, 3, 5\}$, $B = \{1, 3, 5\}$

$P(A) = 3/6 = 1/2$ $P(B) = 3/6 = 1/2$

or

$P(A \cup B) = \{1, 2, 3, 5\} = 4/6 = 2/3$

$1/2 + 1/2 = 1 \neq 2/3 = P(A \cup B)$

∴ for not mutually exclusive events : $P(A \cup B) \neq P(A) + P(B)$

Independent Events

Occurence of one event doesnot affect the occurence of the other event.

$$P(A \cap B) = P(A) \cdot P(B)$$ (and)

occurence of one event doesn't affect the occurence of other.

for independent events

$$P(A \cap B) = P(A) \cdot P(B)$$

Prob : S_1

S_2

$\longrightarrow$ $A = \{S_1 \text{ will pass}\}$, $A^c = \{S_2 \text{ will fail}$

$P(A) = 1/2$

$P(A^c) = 1/2$

$\longrightarrow$ $B = \{S_2 \text{ will pass}\}$, $B^c = \{S_2 \text{ will fail}\}$

$P(B) = 2/3$

$P(B^c) = 1/3$

Probability both fail $\Rightarrow P(A^c \cap B^c) = 1/2 \cdot 1/3 = 1/6$

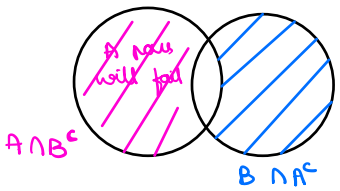
Probability both pass $\Rightarrow P(A \cap B) = 1/2 \cdot 2/3 = 2/6 = 1/3$

exactly one will pass $\longrightarrow$ ① A pass & B fails $= 1/2 \cdot 1/3 = 1/6$

② A fail & B pass $= 1/2 \cdot 2/3 = 2/6 = 1/3$

either ① or ② ie ① ∪ ②

$1/6 + 1/3 = 1/6 + 2/6 = 3/6 = 1/2$ //



$$P\left((A \cap B^c) \cup (B \cap A^c)\right) = P(A \cap B^c) \cup P(A^c \cap B)$$
$$= P(A \cap B^c) + P(A^c \cap B)$$
$$= P(A) \cdot P(B^c) + P(A^c) \cdot P(B)$$
$$= 1/2 \cdot 1/3 + 1/2 \cdot 2/3$$
$$= 1/2$$

Atleast one will pass → one will pass or Both will pass

→ therefore union

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

go through definitions

Solve the following problem

Prob : There are three students

i) all of them will fail

ii) all of them will pass

iii) Exactly one of them will pass

iv) exactly two of them will pass.

soln : i) all fail

S_1 fail, S_2 fail & S_3 fail

Independent events

$\therefore P(S_{1f} \cap S_{2f} \cap S_{3f}) = 1/2 * 1/3 * 1/4 = 1/24$

ii) all will pass

S_1 pass, S_2 pass & S_3 pass

Independent events

$\therefore P(S_{1p} \cap S_{2p} \cap S_{3p}) = 1/2 * 1/3 * 3/4 = 1/4$

iii) Exactly one will pass

S_1 pass & S_2, S_3 fail $= 1/2 * 1/3 * 1/4 = 1/24$

S_2 pass & S_1, S_3 fail $= 2/3 * 1/2 * 1/4 = 1/12$

S_3 pass & S_1, S_2 fail $= 3/4 * 1/3 * 1/2 = 1/2$

$1/24 + 1/12 + 1/2 = 1/24 + 2/24 + 12/24 = 15/24$

iv) Exactly two will pass

$S_{1p}, S_{2p}, S_{3f} = 1/2 * 2/3 * 1/4 = 1/12$

$S_{1p}, S_{3p}, S_{2f} = 1/2 * 3/4 * 1/3 = 1/8$

$S_{2p}, S_{3p}, S_{1f} = 2/3 * 3/4 * 1/2 = 1/4$

$= 1/12 + 1/8 + 1/4$

$= 0.4583$

Matrix algebra

Calculus