

Study of past data to find patterns & insights

Using past information, predict the future outcomes

Data Analysis +

Machine Learning



Data Science

Instagram



How much time is being spent on the platform by the users from different countries / age groups etc.



Reels Feature (IGTV)



A/B Testing

5K

Group A

Users keep seeing IGTV feature in the app

5K

Group B

Users who get app update where the app has Reels feature instead of IGTV

↓
Note down
the time spent

42 mins
average

Till 2010
80% → Structured
20% → Unstructured

Types of Data

Structured
Data
↓
Tabular Form of
Data

feature instead
of I4TV
↓
Note down
the time
spent.
↓
1 hour 15 min
average

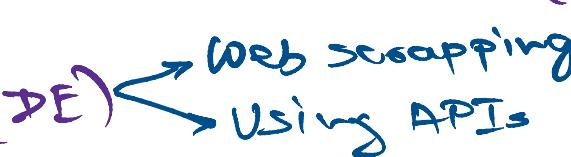
Now
80-90% → Unstructured
10-20% → Structured

Unstructured
Data
↓

Images, videos,
text, audio, signal
etc.

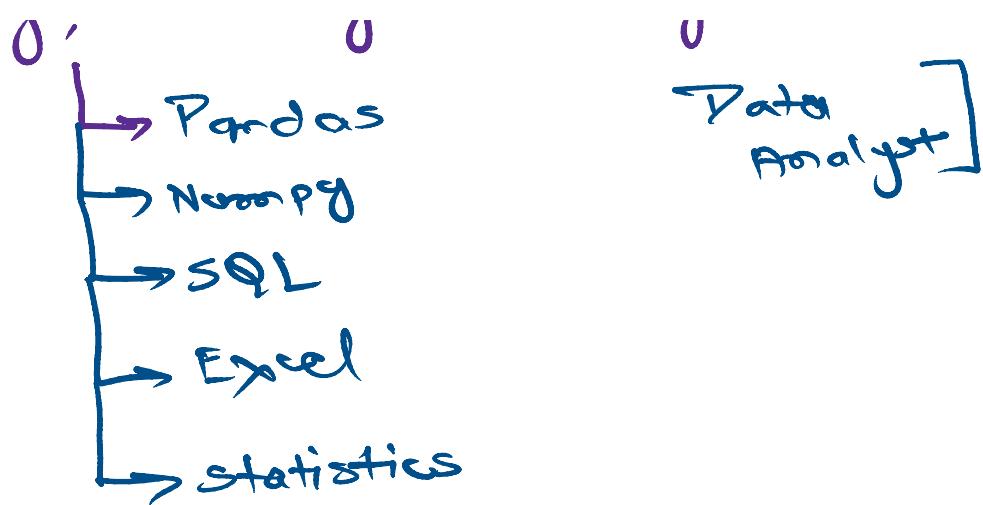
↓
Data Science
(Machine Learning)

Components of Data Science Process →

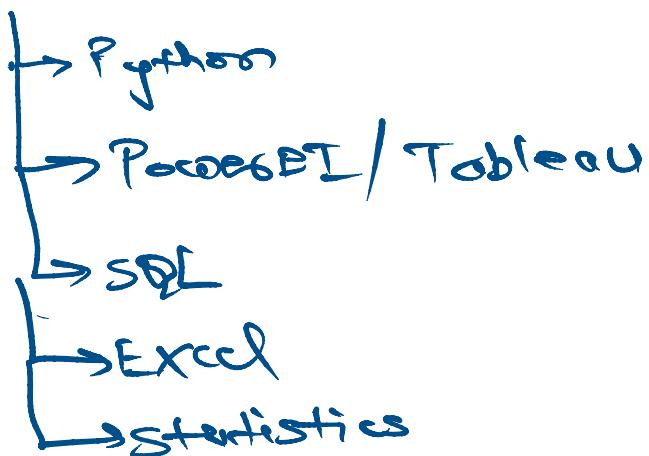
- ① Understanding the problem statement (BA)
- ② Data Collection (DE) 
 - Web Scraping
 - Using APIs
- ③ Processing, structuring & cleaning the data.

I. D. S. C.

Data

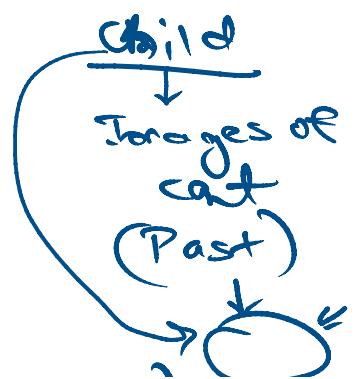


④ Explore & Visualize the Data (DA)



- ⑤ Apply the Machine Learning techniques on the data.
- ⑥ Check the performance of ML model.
- ⑦ Tuning the performance of ML model.
- ⑧ Present the results.

Machine Learning (Child)
+
Training Data (Past)

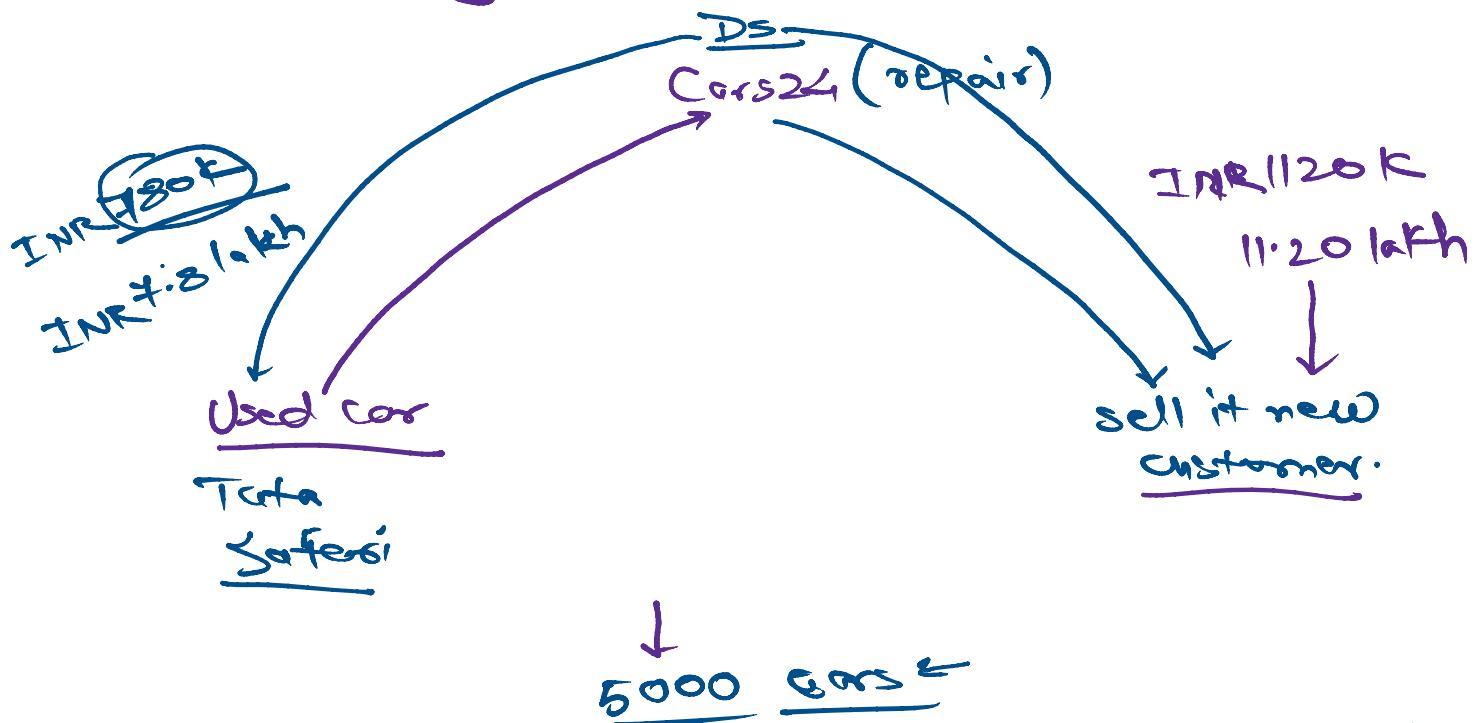


U

↓

Machine Learning Model (Trained)

- ① Make a model for car price prediction for the company 'Cars24'.



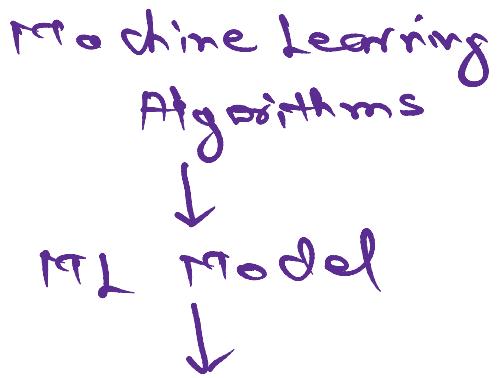
Features in input data

	kms driven	Engine Type	model	Brand	Mileage	mfg Date	Fuel Type	Price
0	2323	Auto	Safari	Tata	16km/l	2020	Petrol	11.210k
1	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-



Machine Learning

Target / output



3427 | Auto | 17.1 km/l | Horison | Tata | 2021 | Diesel → 12.3 lakhs



Equation to represent the relationship
b/w input & output from the given data



To predict Price

Base Price

$$\begin{aligned}
 &= 6.12 + 0.15 \times \text{km's driven} + 0.17 \times \text{engine type} \\
 &+ 0.6 \times \text{Model} + 1.2 \times \text{Brand} + 0.89 \times \\
 &\text{mileage} + 0.51 \times \text{Hg date} \times \\
 &0.28 \times \text{fuel type.}
 \end{aligned}$$

$$= 6.12 + 0.15 \times 3427 + 0.17 \times 1 + \dots$$

↓ ↓

No of Doors Mileage

No of Doors } Mileage

20 future cars

16 correct

4 incorrect

$$\frac{16}{20} = 0.8 / 80\%$$