

## Challenges of ML

### 01. Data collection

API

web scraping

### 02. Insufficient data

03. Non-Representative data (Half data)

### 04. Poor Quality data

### 05. Irrelevant data

06. Over fitting (low bias, high variance)

07. Under fitting (high bias, low variance)

08. Software integration. (IT should be a part of software).

### 09. Offline learning / deployment.

### 10. Cost Involved.

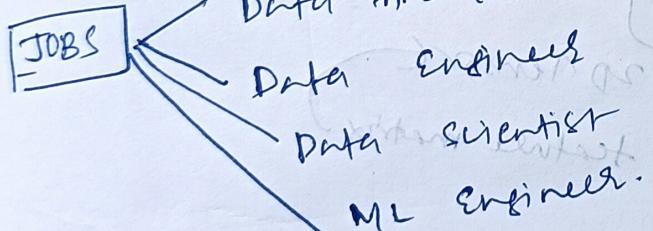
-The cost of AI Research paper.

### Applications [B2B companies]

- ① Retail - Amazon / Big Bazaar
- ② Banking & Finance
- ③ Transport - OLA
- ④ Manufacturing - Tesla
- ⑤ consumer Internet - Twitter

### ML development cycle:

- (i) Frame the problem
- (ii) Data gathering
- (iii) Data preprocessing [Remove duplicates, outliers, missing values]
- (iv) EDA [Exploratory Data Analysis]
  - Visualization
  - Univariate Analysis.
- (v) Feature engineering and selection.
- (vi) Model training, evaluation & selection.
- (vii) Model deployment
- (viii) Testing (A/B testing)
- (ix) Optimize



### Tensor:

- It is a data structure / container for numbers.
- `np.array(4)` → 0D Tensor
- `np.array([1, 2, 3])` → 1D Tensor (Vector)
  - | 2D dimension
- `np.array([1, 2, 3], [4, 5, 6])` → 2D Sensors (Matrix)
  - | (2x3)

$$[[1, 2, 3], [4, 5, 6]]$$

$$[[1, 2, 3], [4, 5, 6]]$$

$$[[1, 2, 3], [4, 5, 6]]$$

## 1D Tensor

| Students (10000) → data |     |       | let $WB = 0$ | Kalnetata = 1 |
|-------------------------|-----|-------|--------------|---------------|
| CUPA                    | 10  | STATE | plummet      | old label     |
| 8.1                     | 911 | WB0   | 1            |               |
| 7.2                     | 002 | 1     | 0            |               |

Features

- consider 1D features  $\{8.1, 911, 0\} \in 1D \text{ Tensor}$

3D Vector

## 2D Tensor

| CUPA      | 10        | STATE     | plummet        | 2D Tensor |
|-----------|-----------|-----------|----------------|-----------|
| [ - - - ] | [ - - - ] | [ Y ]     | → 2D Tensor    |           |
| [ - - - ] | [ - - - ] | [ Y ]     | feature matrix |           |
| [ ; ; ; ] | [ ; ; ; ] | (10000x3) | merging sp.    | 2802      |

## 3D Tensor

NLP — used in NLP (3D Tensor)

| H1 Nitish | H1 | Nitish | Rahul | Amit |
|-----------|----|--------|-------|------|
| H1 Rahul  | 1  | 0      | 0     | 0    |
| H1 Amit   | 0  | 1      | 0     | 0    |

$$\begin{bmatrix} [1, 0, 0], [0, 1, 0] \\ [1, 0, 0], [0, 0, 1] \\ [1, 0, 0], [0, 0, 1] \end{bmatrix} \rightarrow 3D \text{ Tensor} \quad (3 \times 2 \times 3)$$

4D-Tensor - works with images

5D Tensor - works with videos.

Jupyter Notebook, Spyder, Colab  $\Rightarrow$  IDE's for ML

### Create Virtual env

- open anaconda prompt.

01. `conda create --name virtualenvname`

02. `conda activate virtualenvname`

03. `conda install -c anaconda jupyter`

### From kaggle to colab

01. Download the API key token JSON file

02. Write the command

```
!mkdir -p nl.kaggle  
!cp kaggle.json nl.kaggle!
```

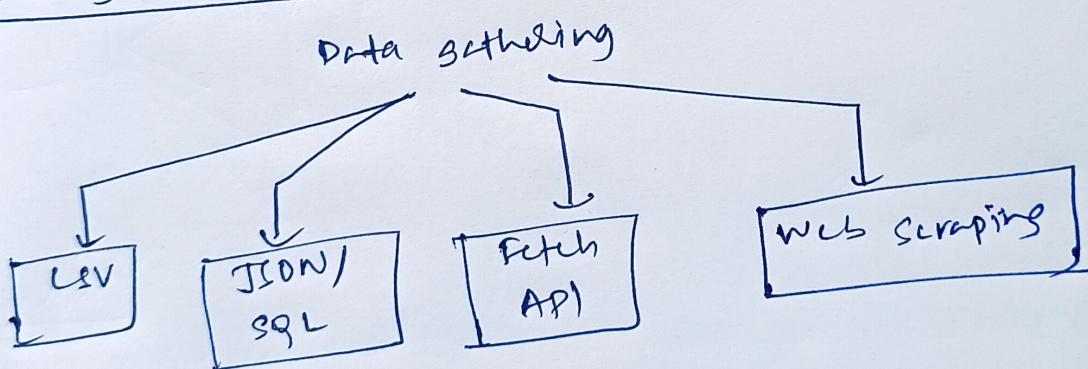
OLTP

03. ! kaggle dataset view

04. Unzip the file

```
import zipfile  
zip_ref = zipfile.ZipFile('zip', 'r')  
zip_ref.extractall('content')  
zip_ref.close()
```

### Working with CSV



```
df = pd.read_csv('cars.csv')
```