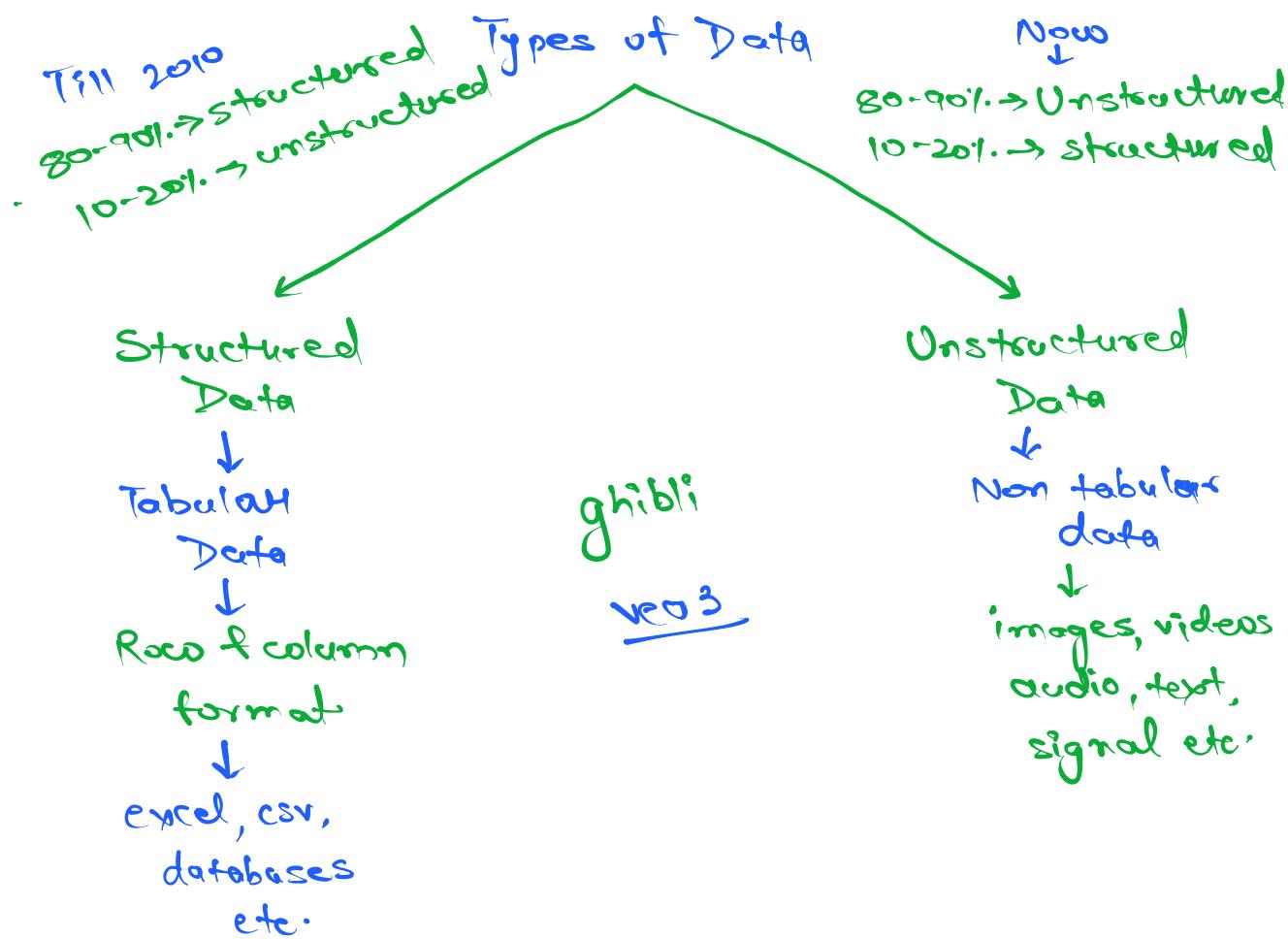


Past Data: Data which we have available with us currently. Of the data which we have already collected.



Machine Learning has two versions →

① Statistical Machine Learning -
↳ Structured Data

② Deep Learning -
↳ Unstructured Data + Structured

Components of a Data Science Project:

① Understanding the business requirements |
problem statements. Budget + Deadline
↳ (Business Analyst)

② Data collection
↳ APIs
Web scraping
servers
surveys

Hadoop
RySpark
Hive
Pig

↳ (Data Engineers)

③ Data Cleaning & Manipulation →
→ Removing missing values, duplicates,
outliers, corrupted data, unnecessary
data etc.
→ Data Wrangling / Modifying the data.

↳ (Data Engineer + Data Analyst)

Tools → NumPy & Pandas

④ Exploratory Data Analysis (EDA)

Tool Req:

- SQL
- Excel
- Statistics
- Tableau / PowerBI
- Python Packages

(Data Analyst)

⑤ Present the analysis report. (DA)

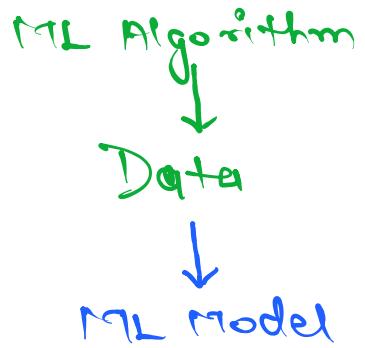
⑥ Data preprocessing for machine learning

- Encoding / vectorization
- Scaling the Data: standardization / Normalization
- Split the data into training & testing sets.

Machine Learning Engineers (MLE)

⑦ Apply Machine Learning Algorithms on the Data.

Yum



Machine Learning Engineers (MLE)

- ⑧ Evaluation & Improvement of the
ML model (MLE)

- ⑨ Machine Learning Operations (MLOps)

Continuous Development &

Continuous improvement

(CI/CD Pipeline)

&

Deployment of ML model in the
cloud

(AWS, Azure, GCP, Oracle Cloud)

(MLE)

Machine Learning Process →

Cars24

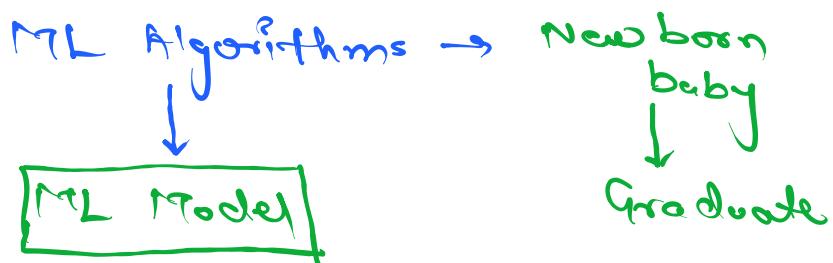
New owner

Price/output

features/input

Lakh
1.21 lakh
INR 120k

B	C	D	E	F	G	H	I	J	K
selling_price	year	seller_type	km_driven	fuel_type	transmission	mileage	engine	max_power	seats
1.2	2012	Individual	120000	Petrol	Manual	19.7	796	46.3	5
5.5	2016	Individual	20000	Petrol	Manual	18.9	1197	82	5
2.15	2010	Individual	60000	Petrol	Manual	17	1197	80	5
2.26	2012	Individual	37000	Petrol	Manual	20.92	998	67.1	5
5.7	2015	Dealer	30000	Diesel	Manual	22.77	1498	98.59	5
3.5	2013	Individual	35000	Petrol	Manual	18.9	998	67.1	5
3.15	2013	Dealer	40000	Petrol	Manual	20.36	1197	78.9	5
4.1	2018	Dealer	17512	Petrol	Manual	20.51	998	67.04	5
10.5	2019	Individual	20000	Petrol	Automatic	18.15	998	118.35	5
5.75	2017	Dealer	70000	Diesel	Manual	18.49	1493	100	7
3.05	2015	Individual	50000	Diesel	Manual	19.09	1405	69.01	5
11.5	2019	Individual	18000	Diesel	Manual	20.37	1461	108.49	5
5.11	2017	Dealer	28321	Petrol	Manual	16.6	1197	85	5
4.1	2016	Dealer	27000	Petrol	Automatic	19.34	1198	76	5



ML Model

$$y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + c$$

weights ↓

$$\rightarrow \left[\begin{array}{l} \text{Price} = 0.87 \times \text{Mileage} + 0.91 \times \text{Engine Power} + \\ 1.21 \times \text{Brand} + (-1.05) \times \text{Kms driven} + \end{array} \right]$$

L 1.09 → Base Price J

New car: 17.1 km/l 990 Total 15672