

Study of past data → Data Analysis
to find observation & insights.

+
Machine Learning → Predict the future outcomes on the basis of past data.
↓
Data Science

80-90's → Structured
10-20's → Unstructured
Till 2011

Types of Data

Structured Data
→ Tabular Format (Rows & cols)

Excel files,
databases
etc

Excel, SQL
etc

~~DS / ML~~
Solve a business problem.

10-20's → Structured
80-90's → Unstructured

Now

Unstructured Data.

→ Images, videos,
text, emojis,
audio, signal etc

Mongodb,
Cassandra

Machine Learning / DS
1950

2014
512 MB
256 MB
1 GB

Lack of
Data

Limited
computational
power.

2013-14

Components of a Data Science Process →

- ① Understanding the problem statement & pain points of the business.

(Business Analyst) →

Budget + Deadline.

- ② Collect the Data → Data Engineer.

- ③ Data Cleaning & Manipulation

→ Removing missing values, duplicates, outliers, corrupted data, unnecessary data.

→ Changing the available data.

(Data Engineer + Data Analyst)

Tools Req: NumPy & Pandas, Excel, SQL etc.

④ Exploratory Data Analysis (EDA) → Data Analyst

Tools :

- SQL / NoSQL
- Python / R
- Tableau / Power BI
- Excel
- Statistics

⑤ Share the analysis report.

⑥ Data Preprocessing for Machine Learning.

Feature
Engineering

- Encoding the categorical data.
- Scaling the Data.
- Split the data into training & testing.

Machine Learning Engineer.

⑦ Apply ML Techniques on the Data.

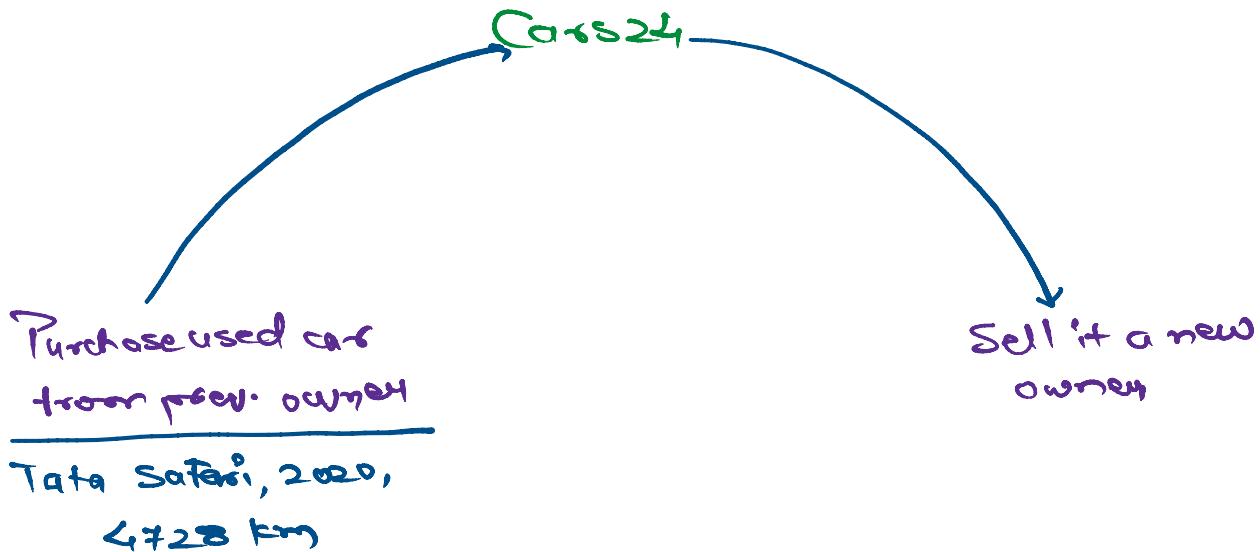
(MLE)

⑧ Machine Learning Operations (MLOps)

Continuous Development &
Maintenance of ML Model
+
(MLE)

Deploy the model in Cloud
(AWS, Azure, GCP)

✓ Example of ML Process



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
2.05	2015	Individual	50000	Diesel	Manual	19.00	1105	60.04	5

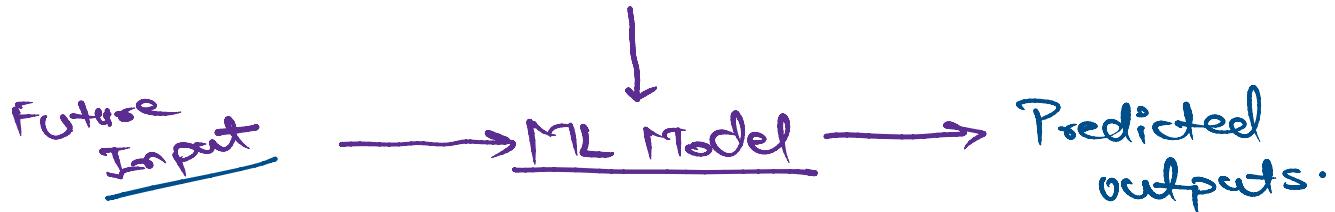
B	C	D	E	F	G	H	I	J	K
selling_price	year	seller_type	km_driven	fuel_type	transmission	mileage	engine	max_power	seats
1.2	2012	Individual	120000	Petrol	Manual	19.7	796	46.3	5
5.5	2016	Individual	20000	Petrol	Manual	18.9	1197	82	5
2.15	2010	Individual	60000	Petrol	Manual	17	1197	80	5
2.26	2012	Individual	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



Machine Learning Algorithms



Learn which factors were affected the price of the car and by what amount.



Past Data



Data we have

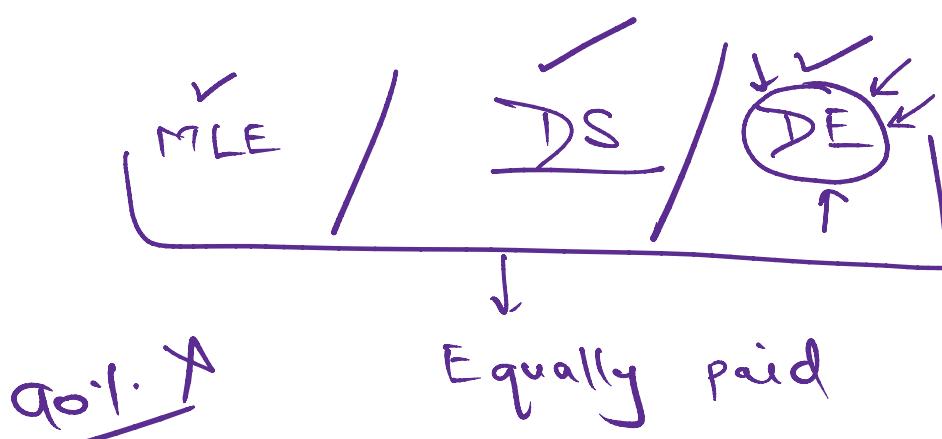
2018 → 3 April 2025

2018 → 2020

Google searches

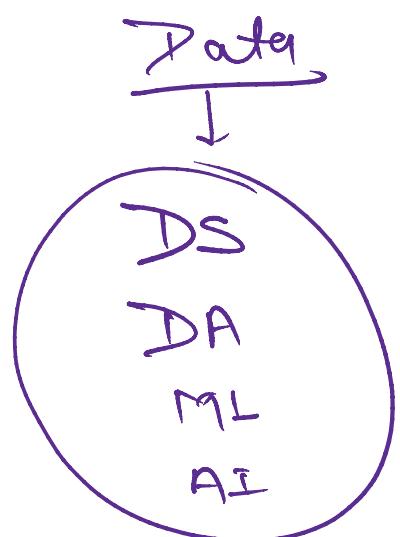
Data processing
↓
Data Analysis

↑
ML
↓
Encoding



go-l. ↑

1.1. → 100.1.
10.1. 100.1.

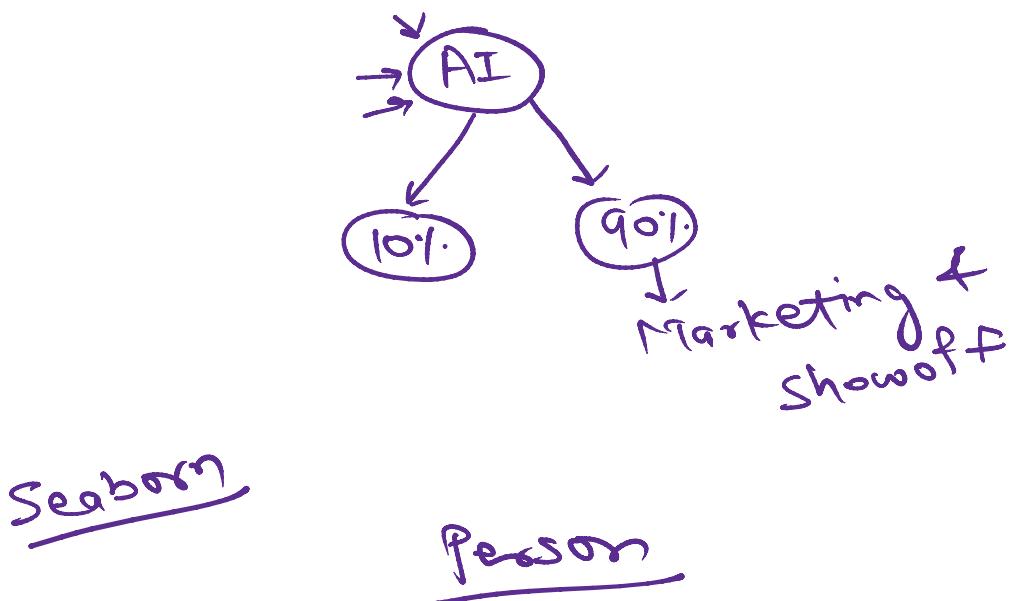


DS
↓

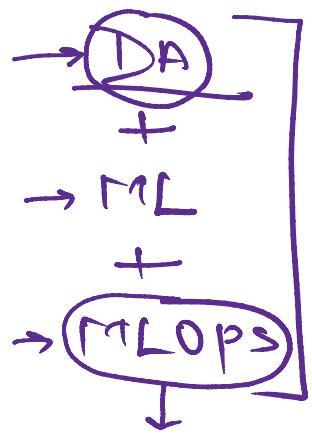
Business Analyst

DS →

- Business Analyst
- Data Analyst
- Research Analyst
- Research Scientist
- Machine Learning Eng.
- MLOps Eng.



→ Links Torvalds
 ↓
 Linux, Git



Linear Regression
Logistic Regression

3 lakh → 40 Lakh