Introduction to Machine Learning 26 Jan

26 January 2025 09:57

We study the past data to find observations & insights. → Data Analysis

+

Machine Learning → Predict the future outcomes on the basis of past data.

↓

Data Science

Types of Data

Till 2010

80-90% → structured

10-20% → Unstructured

↓

Structured data

↓

Tabular Form (Rows & cols)

Excel files, databases etc

↓

Excel, SQL etc

Now

10-20% → structured

80-90% → Unstructured

↓

Unstructured Data

↓

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

↓

Machine Learning/ DS

HDFC Bank

1-2 years

1
2
3
4
⋮
1000

Loan → HDFC → 1 — Rej, 2 — Acc, 3 — Rej, 4 — Rej

7-8 years

Previous loans applications which were accepted or rejected

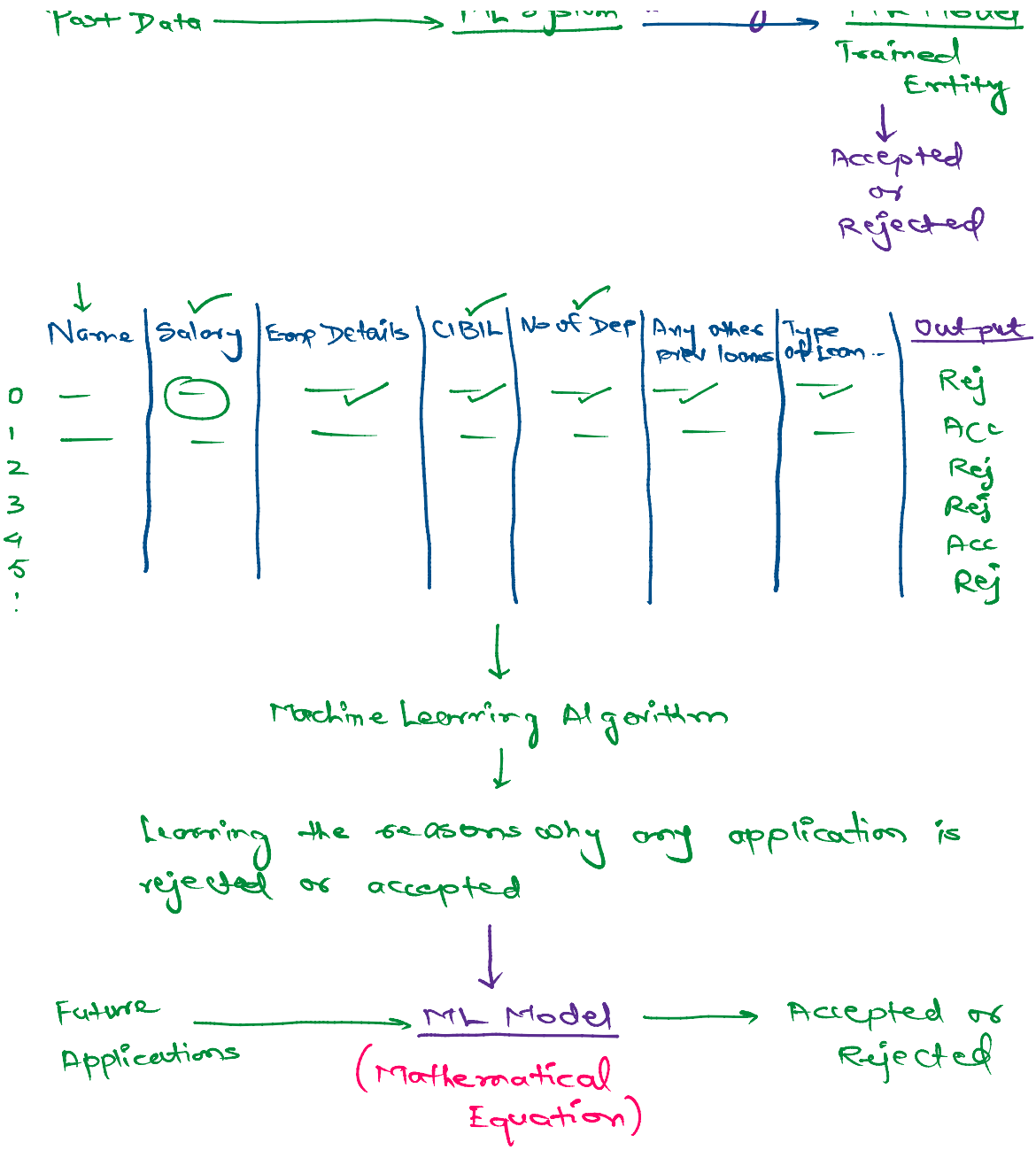
↓

Past Data → ML system → Learning → ML Model

Trained Entity

New Applications

↓



Components of a Data Science Process →

1. Understanding the Problem statement & pain points.
   (Business Analyst)
2. Data Collection
   - APIs
   - Web scraping
   - Server
   - Surveys

   ↓ Data Engineer
3. Data Cleaning & Manipulation

③ Data Cleaning & Manipulation

- → Removing missing values, duplicates, outliers, corrupted data, unnecessary data, etc.
- → Modifying the data.

(Data Engineer + Data Analyst)

Tools Req: Numpy, Pandas

④ Exploratory Data Analysis (EDA) → Data Analyst

Tools Req: SQL

- → Python / R
- → Tableau / PowerBI
- → Excel
- → Statistics

⑤ Data Preprocessing for Machine Learning

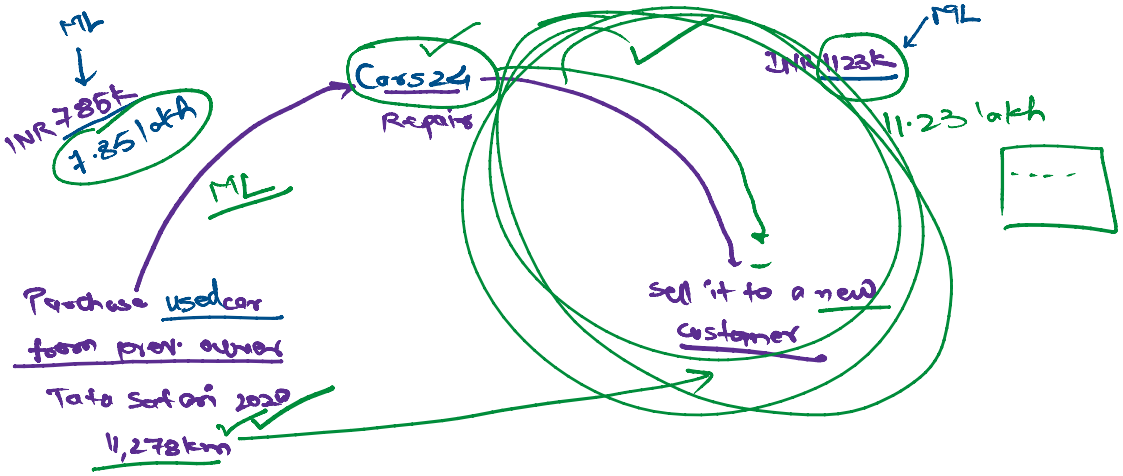
- → Encoding the categorical data.
- → Scaling the data: Standardization / Normalization
- → Split the data into training & testing set.

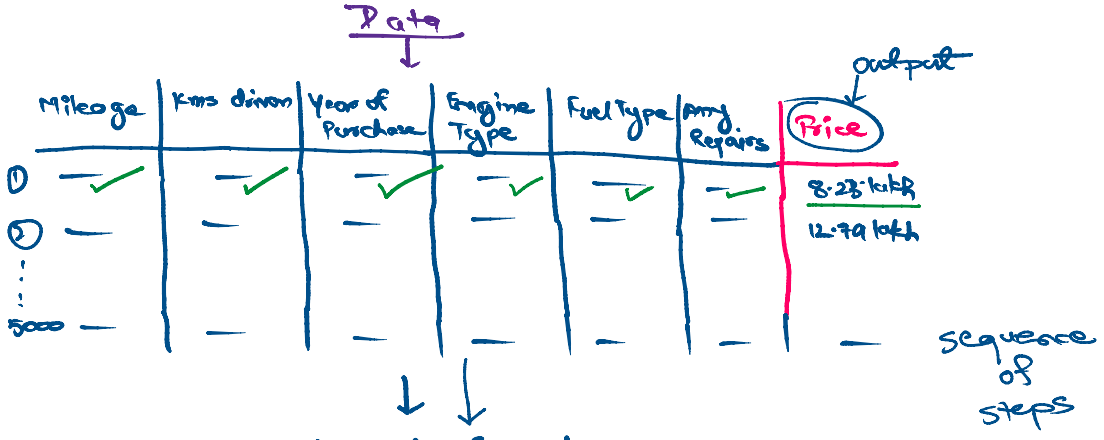
Machine Learning Engineer

⑥ Apply Machine Learning on the data. (MLE)

⑦ Deploy on cloud (Azure, AWS) DS Project

Another ML Process Example





ML Algorithm / system

Learn the relationship b/w input features & price of the car.

ML Model

17 km/l	9872 km	2021	Manual	Petrol	No	→ 13.77 lakh

INR 377k

4.38 lakh

Base Price

weightage

Predicted Price = 4.38 + 0.98 × Mileage + 0.17 × kms driven +

0.6 × Year of Purchase + [illegible]

Price

0.6 × year of Purchase + 0.22 × Engine Type +

0.58 × Fuel Type + (-1.21) × repairs