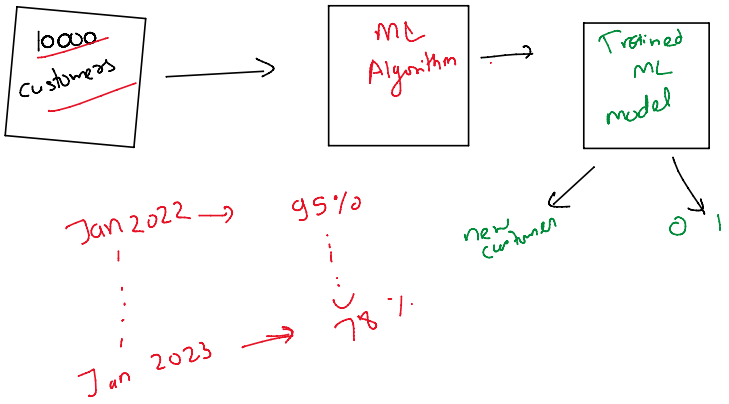
Introduction to ML

27 January 2022 09:22

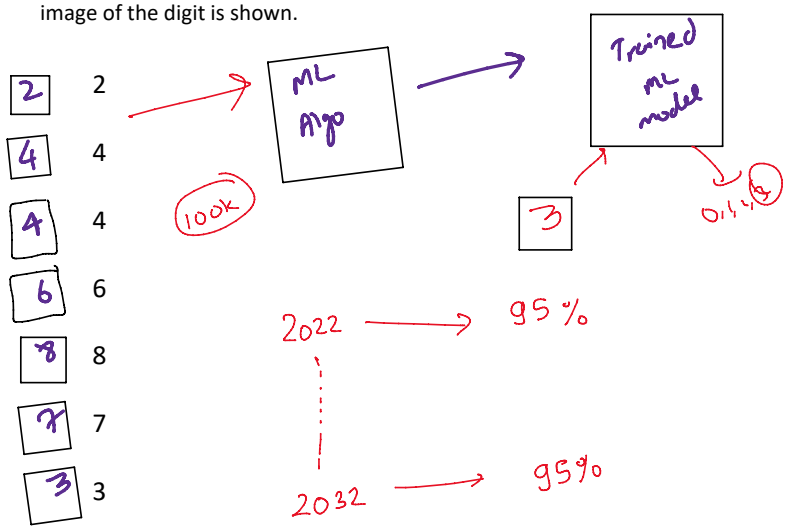
Supervised Machine Learning

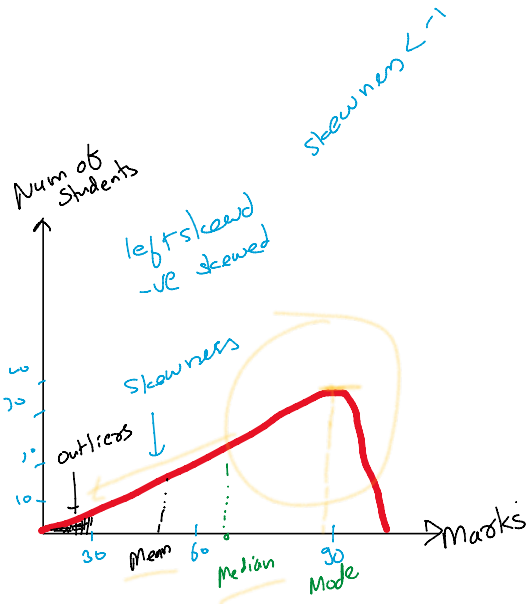
- Used to solve problems where we want to make predictions
- Always need labelled business data
- Supervised ML algorithms learn from data, a trained supervised model does not carry the training data, it carries the patterns in some mathematical format.
- A trained supervised model when deployed to production, it does not learn in production, it stays static in production. It only takes real data and makes predictions in production.

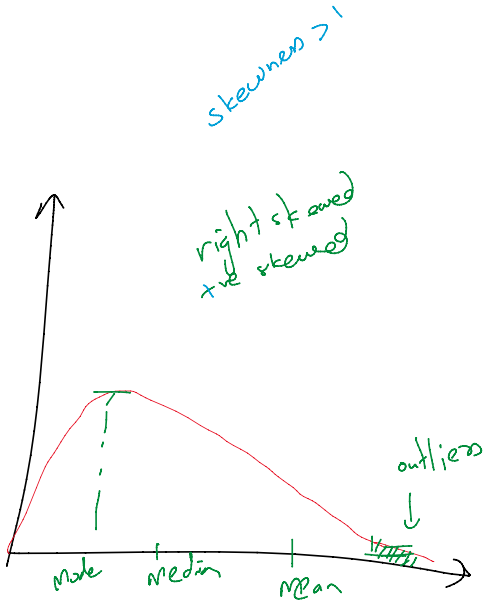
Case 1 : Banking customer churn prediction
Build an ML solution to predict whether customer will leave or will not leave the bank.

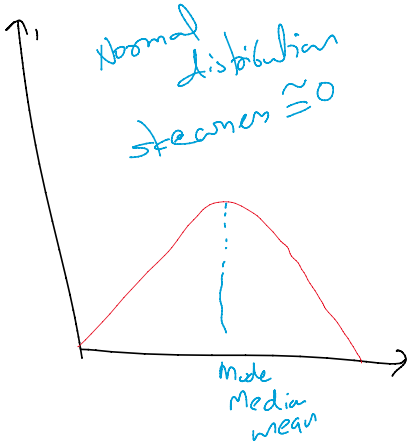


Case 2 : Banking cheque digit identification
Build an ML model which can read the digit when an image of the digit is shown.









Accuracy

	Actual	Prediction
1	0	0
2	0	0
3	1	0
.	0	1
.	1	1
.	1	1
7	1	0
8	1	0
9	1	0
.	0	0
.	0	1
.	1	0
13	1	0
14	0	0
15	0	0
.	0	0
.	0	0
.	0	0
19	0	0
20	1	0

Accuracy = 11/20 = 0.55
Recall = 2/9=0.2222
Precision = 2/4 = 0.5

$$f1\ score = \frac{2}{\frac{1}{R} + \frac{1}{P}}$$

Performance analysis
Test data - 100 customers
Acutal
90 - 0 - not leave
10 - 1 - left

Prediction
95 - 0 - not leave - 90
05 - 1 - left - 05

Accuracy = 95/100 = 0.95
Relevant accuracy (recall - accuracy in class 1) = 5/10 = 0.5

Prediction - model2
80 - 0 - not leave - 80
20 - 1 - leave - 10

Accuracy = 90/100 = 0.9
Recall = 10/10 = 1

Journey of any organization with data

1. Step 1
   a. Question - what has happened in past?, what is happening now?
   b. These questions are usually permanent and continuous.
   c. Descriptive Analytics - Business Intelligence
   d. We build dashboards which show important KPIs answering business questions
   e. Tools: Power BI / Tableau
   f. Technology: Data Visualization, Statistics

2. Step 2
   a. Question - why it has happened in the past?
   b. These questions are occasional, may occur less often.

c. Exploratory Analytics / Root Cause Analytics
d. Based on need, we perform detailed EDA and create a report
e. Tools: Python, R, Hadoop, spark
f. Technology: Data Visualization, Statistics

3. Step 3
   a. Question: What will happen tomorrow?
   b. These question rise with growth of business, once raised, these question become permanent need of the business.
   c. Predictive Analytics - Predictive Modelling
   d. We build predictive models and deploy in production
   e. Tools: Python, R, spark
   f. Technology: Supervised Machine Learning algorithm

4. Step 4
   a. Questions: How can we make it happen?
   b. These questions are more aspiring ones, but less realizable with less data
   c. Prescriptive Analytics
   d. We build prescriptive models
   e. Tools: Spark, hadoop
   f. Technology: Unsupervised ML