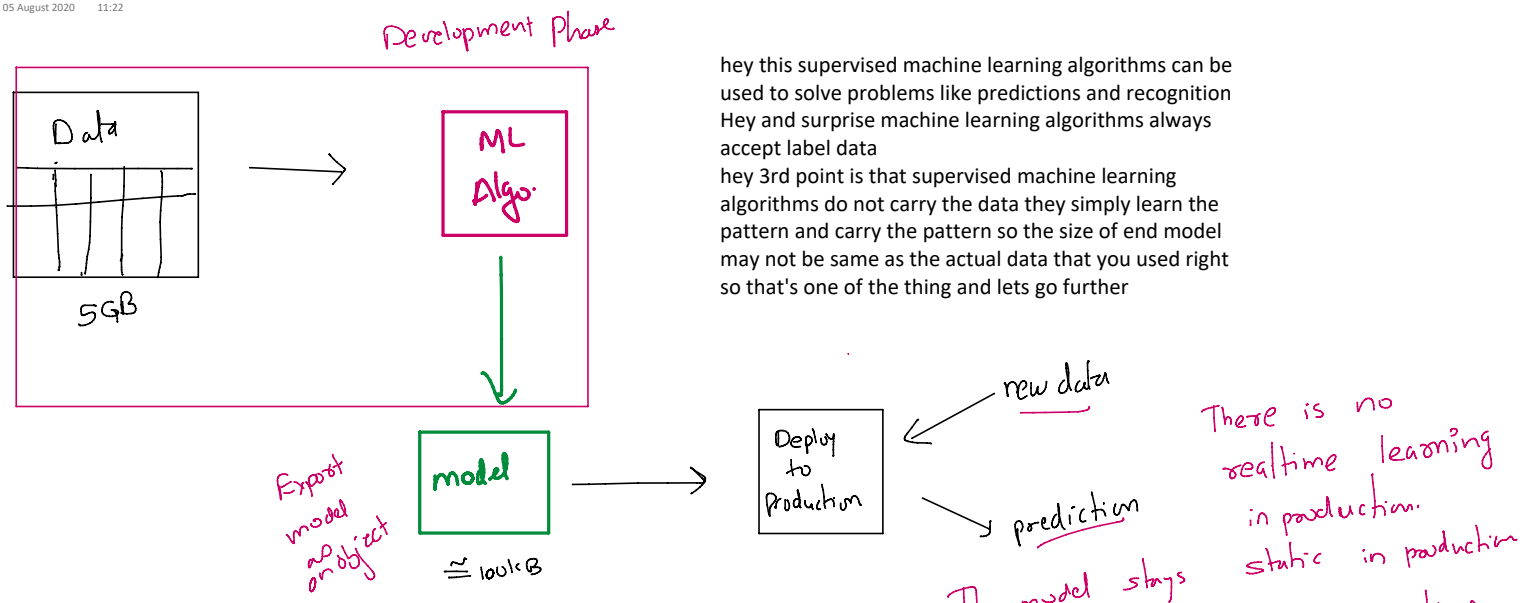
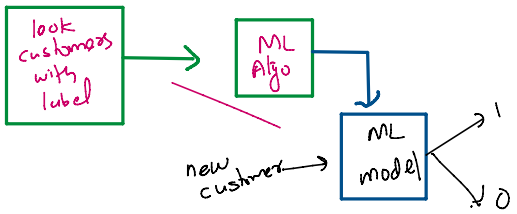
Module 1

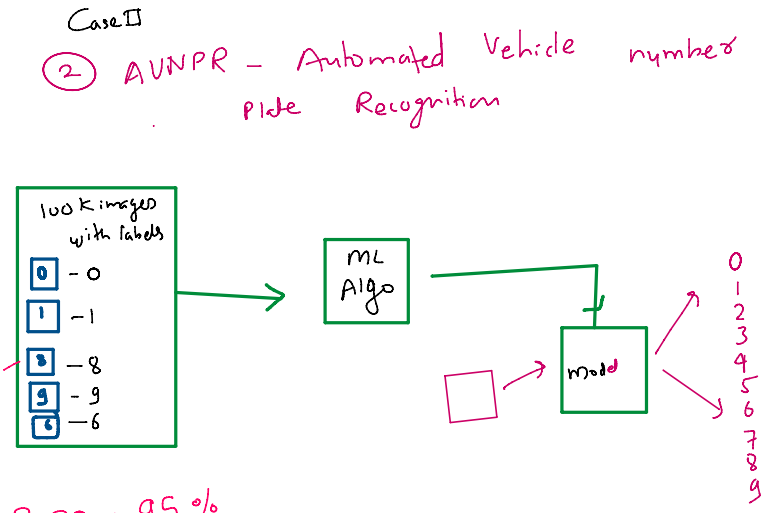
05 August 2020 11:22



hey this supervised machine learning algorithms can be used to solve problems like predictions and recognition
Hey and surprise machine learning algorithms always accept label data
hey 3rd point is that supervised machine learning algorithms do not carry the data they simply learn the pattern and carry the pattern so the size of end model may not be same as the actual data that you used right so that's one of the thing and lets go further



August -2020 – 92%
⋮
August 2021 – 80%



2020 – 95%
↑
2030 – 95%

deploy → forget

Journey of any enterprise with Data Science

① What happened in business? / What is happening in business?

Descriptive Analytics

Data Reporting / Business Intelligence — power bi / tableau

② Why this happened in past?

Exploratory Analytics

— Data Analytics — Python / R / Big data tools

③ What will happen tomorrow?

Predictive Analytics

— Supervised Machine Learning - Python / R

④ How can we make it happen?

Prescriptive Analytics

— Unsupervised ML — Python / R / Big data tools

(A) 10, 12, 24, 26, 28, 24, 26, 32, 34, 36, 181, 24, 28

(B) 5, 6, 7, 6, 5, 5, 6, 6, 7, 16, 5, 99

Statistics - 1

Kualalumpur

18
19
18.5
18.1
18.3
18.8
19.1
19.4
19.6
18.4
19.3
19.6
19.4
18.8

$\bar{x} = 19$

low variance

Singapore

11
11.5
12.6
14.8
16.9
18.4
21.1
22.2
24.6
21.8
20.5
16.3
18.1
16.4
18.3
14.1

$\bar{x} = 24$

high variance

Variance $= \sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$

Telecom Sites

Site ID	Temp	Power	Signal strength	Site shutdown
L1K2				1
L1K4				0
SDK1				0
SDK2				0
				1
				1
				0
				0
				0

σ^2(temp) = 0.0004

Standard deviation $= \sqrt{\text{variance}}$

$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$

1000 ± 160

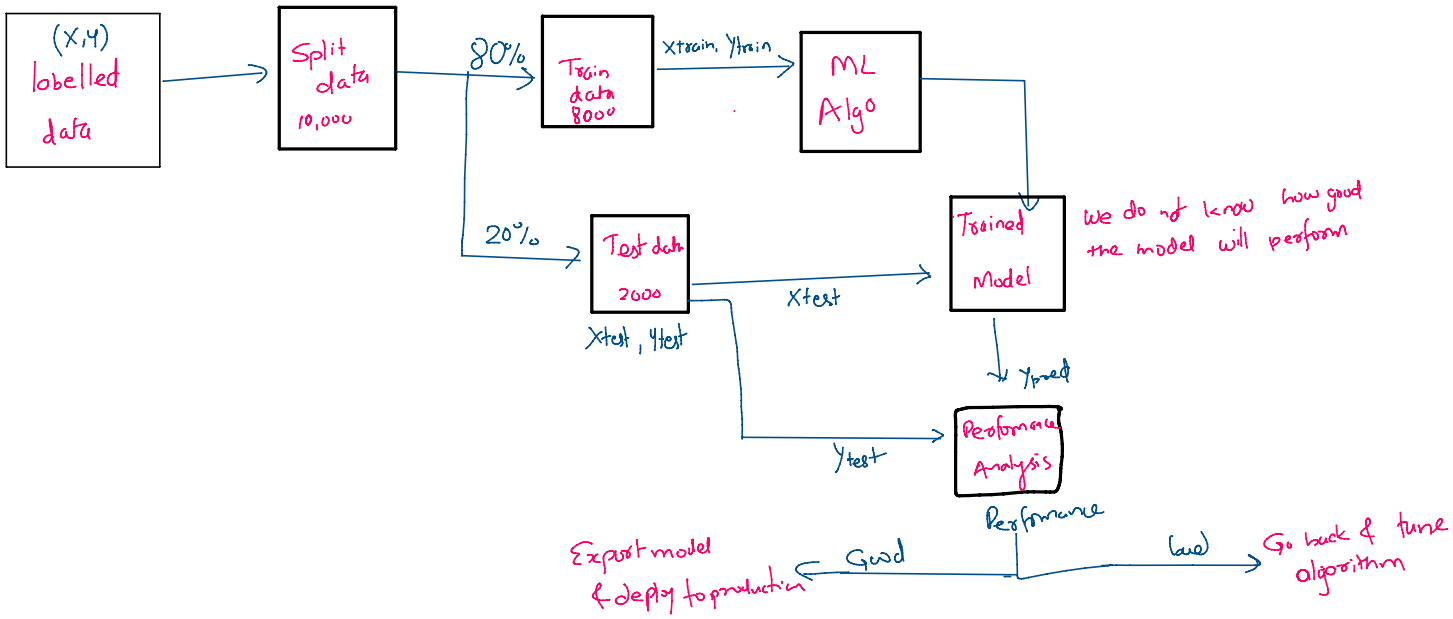
↳ 840 to 1160

	A	B	C
	$1000	$100	$100
$\bar{x}$	$320	$80	$40

Empirical rule

95% $\bar{x} \pm 2\sigma$

Machine Learning Lifecycle



Business Problem Formulation

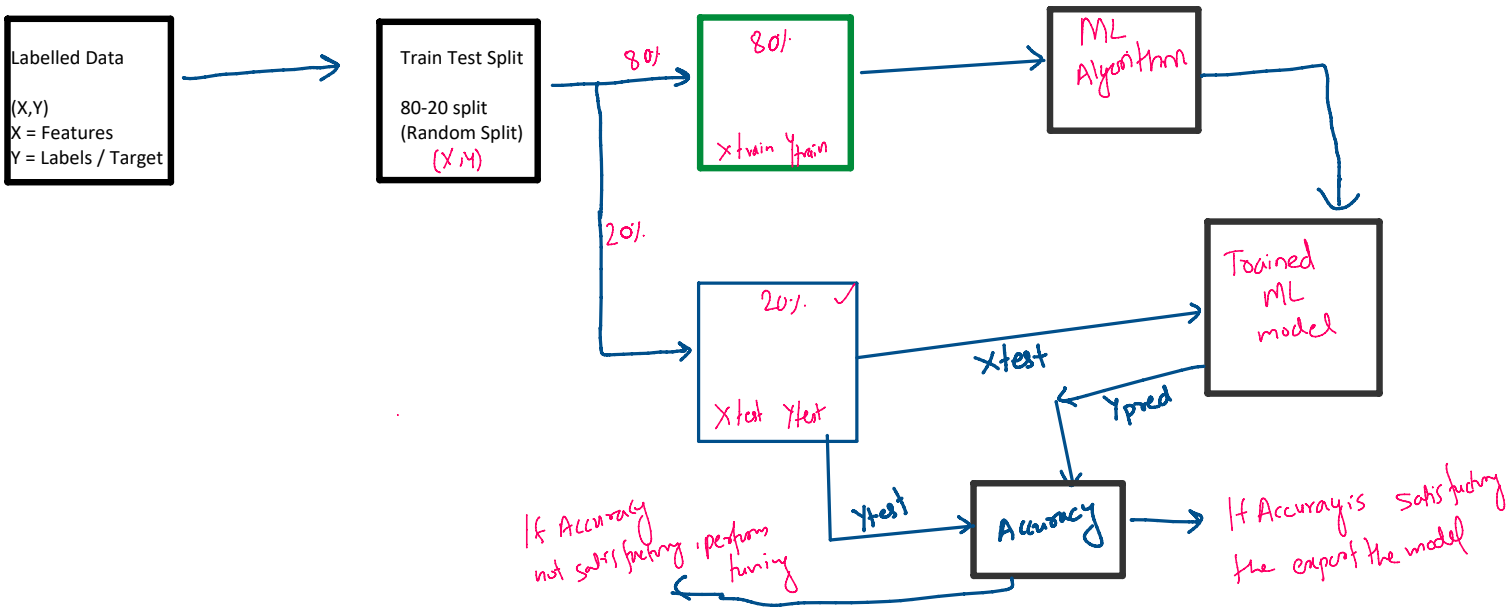
A. Business is struggling with the problem of investment they should make on talent improvement v/s the amount of profit they are making from business. The business wants to have a possible estimation of expected profit they will be able to make in near future so that they can take appropriate decisions.

B. The Product development team wants to design new telecom calling and internet plans, they want the new plans to be more suitable to customers needs and their usage behavior, how can they know about the customer behavior?

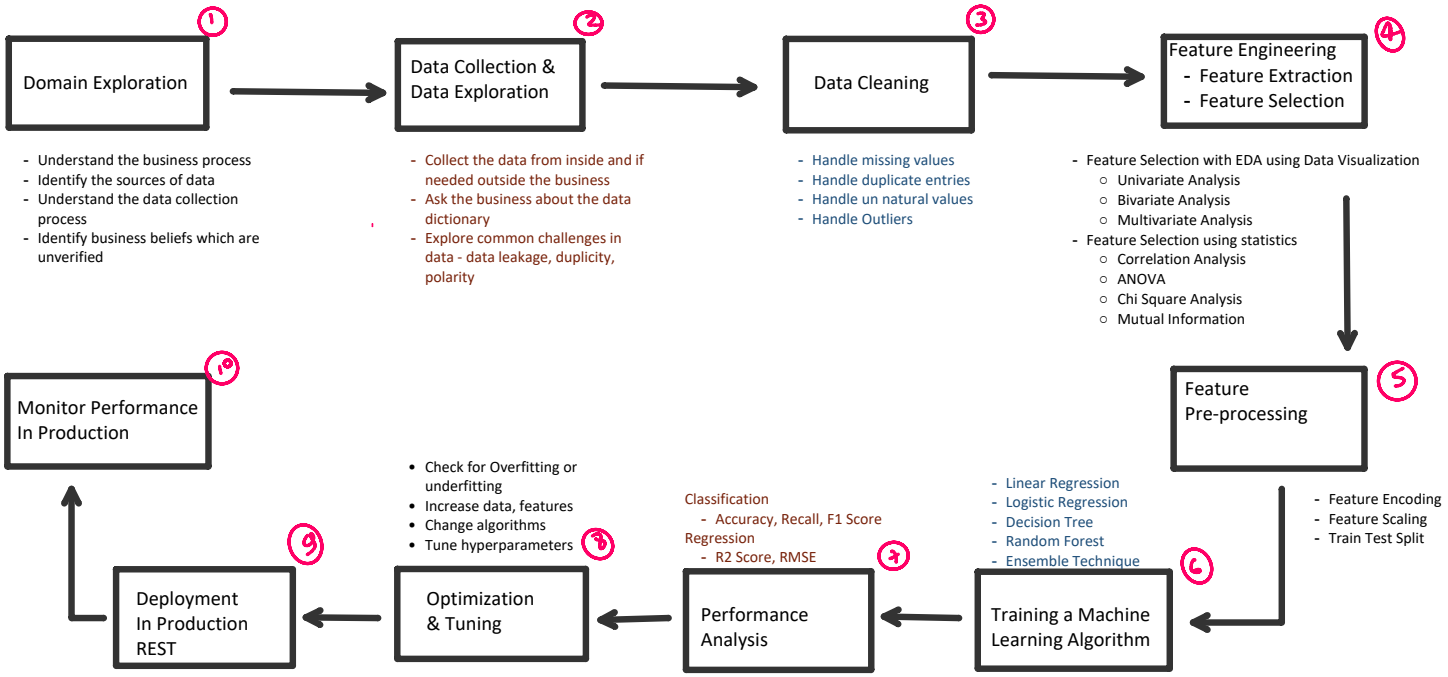
C. The business has more than 1000 calling plans, the customers get confused about which plan they should actually use based on their needs.

D. The business identified that there are many telecom sites which unexpectedly shut down, they are much struggling with unexpected shutdown, How can we help them?

Machine Learning Project Life Cycle



Data Science Project Lifecycle



Skewness = $\frac{\bar{x} - Md}{\sigma}$ = -ve

92+95+98+97+94+96+92+92+92+92+91+25+24+22+23 = 1,125
Mean = 1125/15 = 75
Mode = 92
Median = (22,23,24,25,91,92,92,92,92,92,94,95,96,97,98) = 92