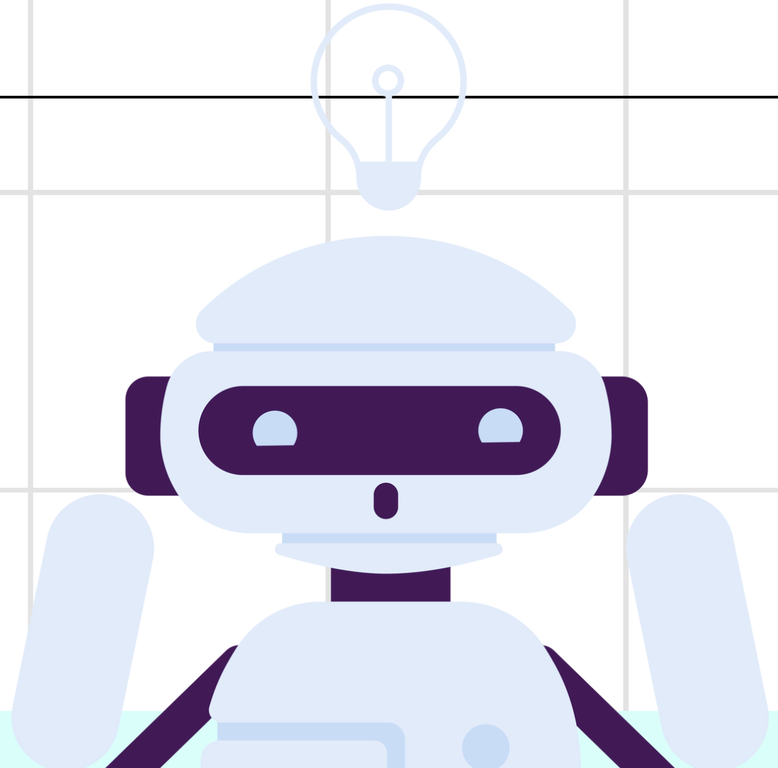
What is....

Supervised Learning?

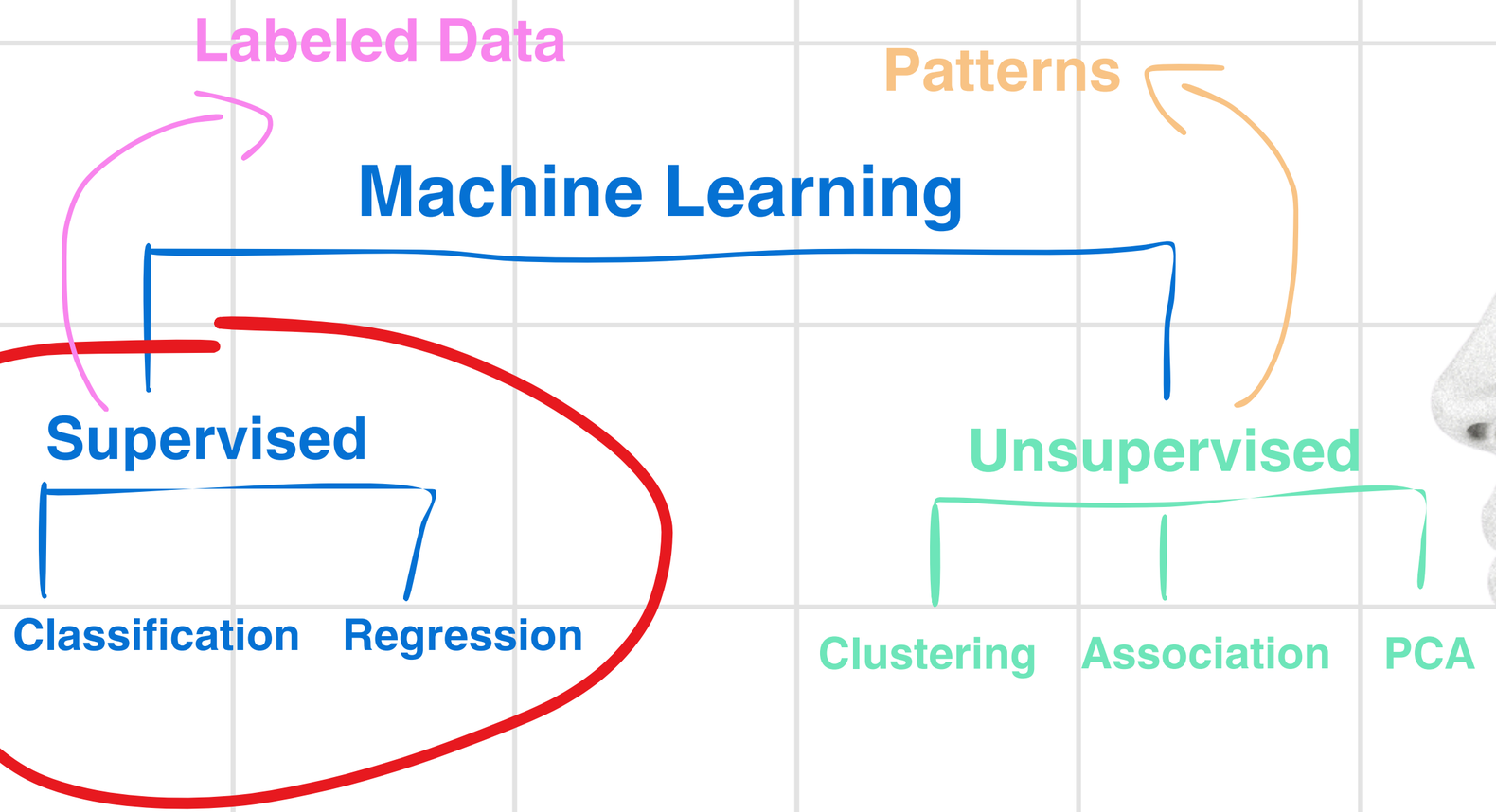
Selma MANI

Link to the notebook we will be using for this session ^.^

Supervised Learning Notebook

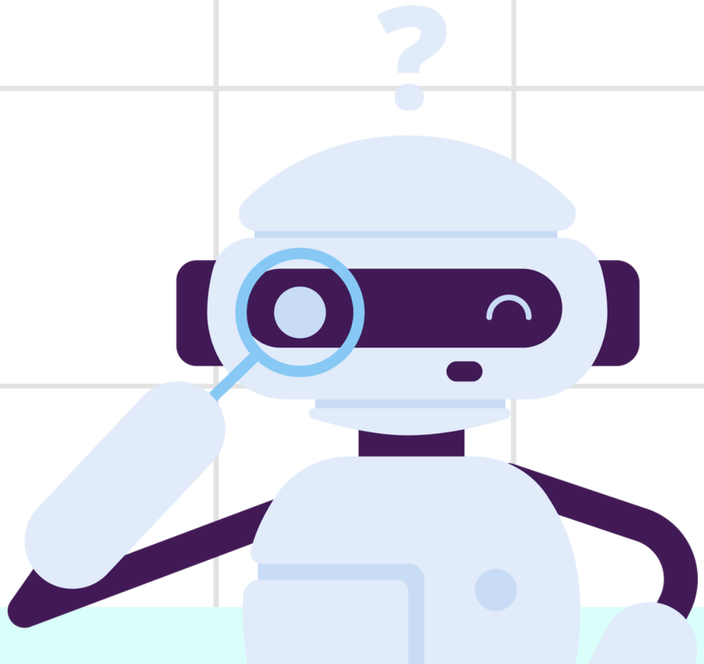


First Let's Define Supervised & Unsupervised Learning...



First Let's Define Supervised & Unsupervised Learning...

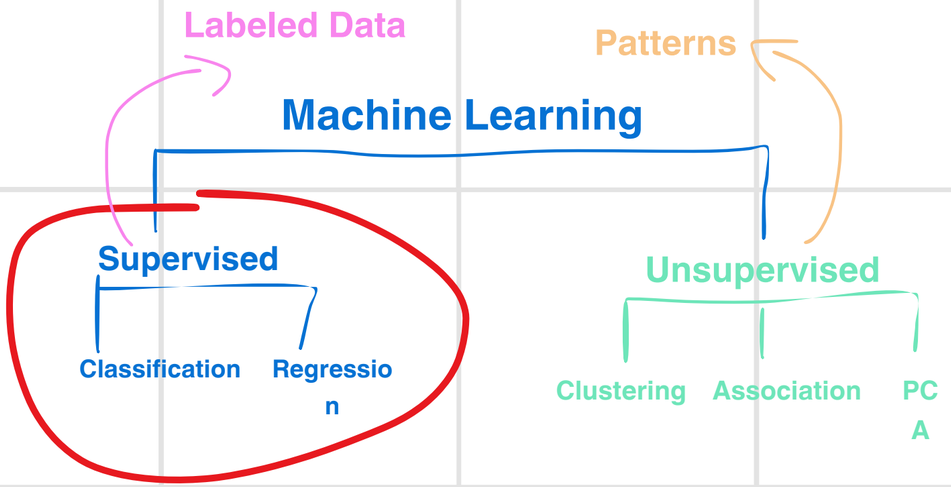
Test Alert!!



	Member_number	Date	itemDescription	year	month	day	day_of_week
0	1808	2015-07-21	tropical fruit	2015	7	21	1
1	2552	2015-05-01	whole milk	2015	5	1	4
2	2300	2015-09-19	pip fruit	2015	9	19	5
3	1187	2015-12-12	other vegetables	2015	12	12	5
4	3037	2015-01-02	whole milk	2015	1	2	4

	Survived	Pclass	Sex	Age	Fare	Embarked	Title	FamilySize
0	0	3	0	22.0	7.2500	0	1	2
1	1	1	1	38.0	71.2833	1	3	2
2	1	3	1	26.0	7.9250	0	2	1
3	1	1	1	35.0	53.1000	0	3	2
4	0	3	0	35.0	8.0500	0	1	1

First Let's Define Supervised & Unsupervised Learning...



Transactions

	Member_number	Date	itemDescription	year	month	day	day_of_week
0	1808	2015-07-21	tropical fruit	2015	7	21	1
1	2552	2015-05-01	whole milk	2015	5	1	4
2	2300	2015-09-19	pip fruit	2015	9	19	5
3	1187	2015-12-12	other vegetables	2015	12	12	5
4	3037	2015-01-02	whole milk	2015	1	2	4

Label

	Survived	Pclass	Sex	Age	Fare	Embarked	Title	FamilySize
0	0	3	0	22.0	7.2500	0	1	2
1	1	1	1	38.0	71.2833	1	3	2
2	1	3	1	26.0	7.9250	0	2	1
3	1	1	1	35.0	53.1000	0	3	2
4	0	3	0	35.0	8.0500	0	1	1

Types of Supervised Learning

Classification

Is where the target feature is categorical. The goal is to assign input data to one of several predefined categories. Often, these categories are binary (e.g., 0 or 1), but they can also include multiple labels such as 'spam' or 'ham'.

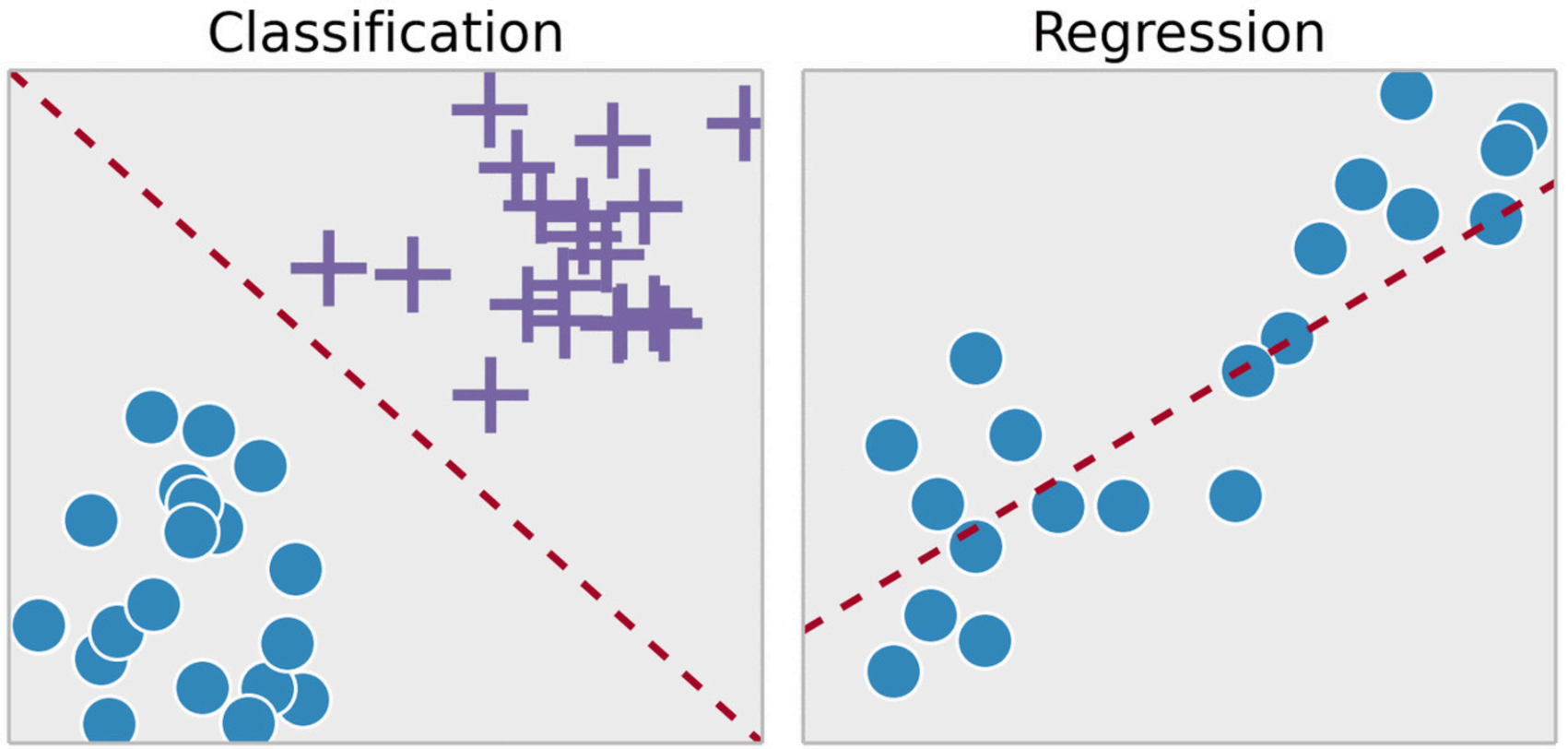
	Outlook	Temperature	Humidity	Wind	t
0	Sunny	Hot	High	Weak	No
1	Sunny	Hot	High	Strong	No
2	Overcast	Hot	High	Weak	Yes
3	Rain	Mild	High	Weak	Yes
4	Rain	Cool	Normal	Weak	Yes

Regression

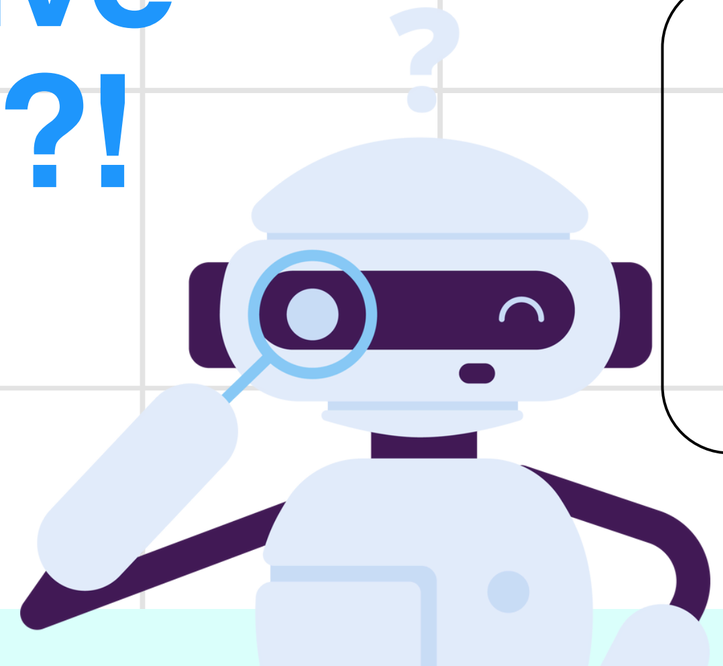
Is where the target feature is continuous. The goal is to predict a numerical value based on input data. Common examples include predicting prices, temperatures, or other measurable quantities.

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

Types of Supervised Learning



Target, Descriptive Features?!



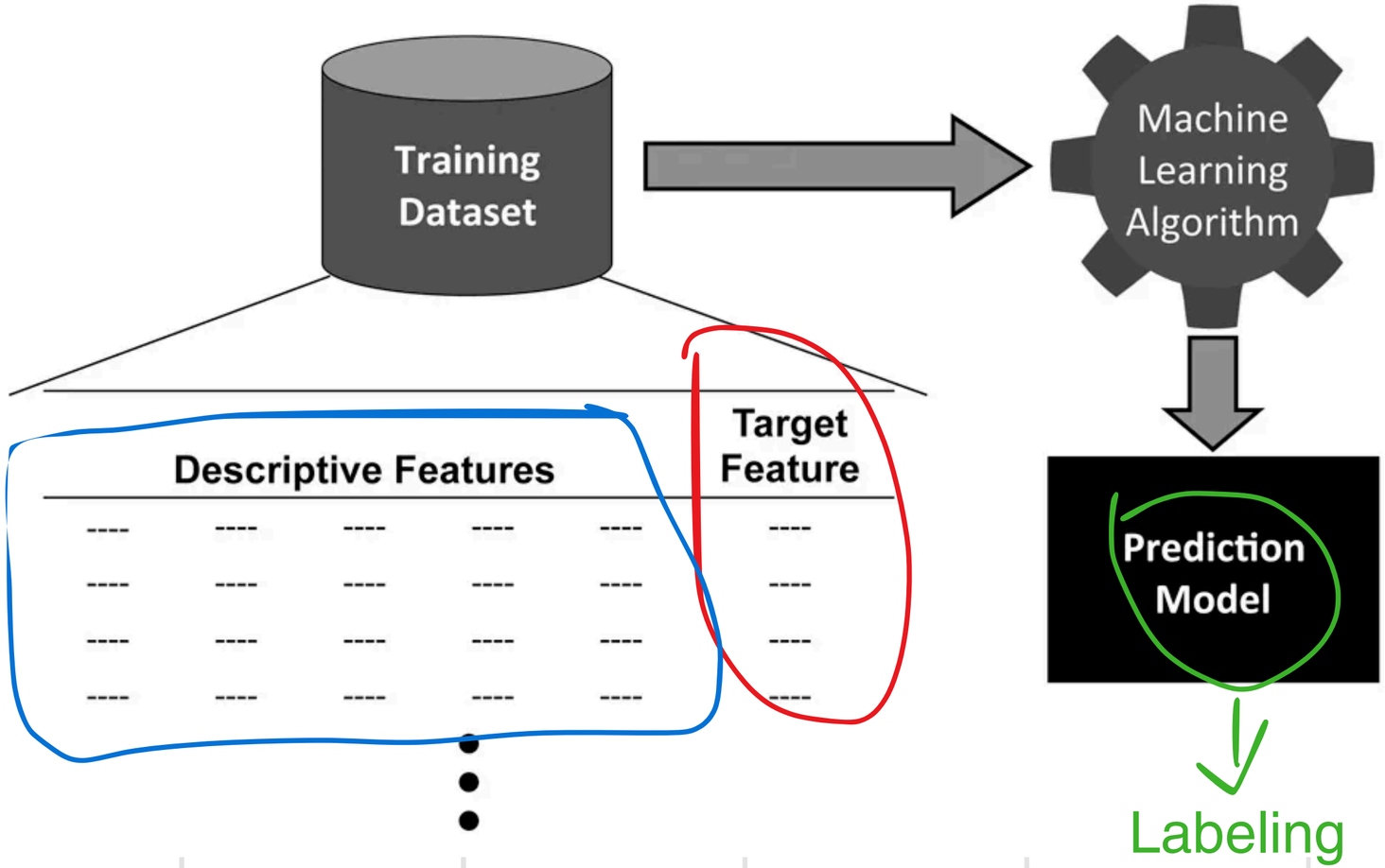
Target Feature

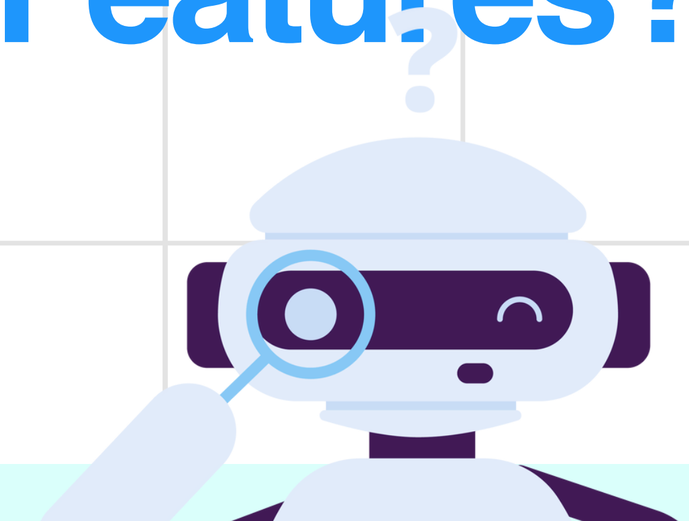
is the variable that you want to predict or classify. It's also known as the dependent variable. For example: weather, salary, etc.

Descriptive Features

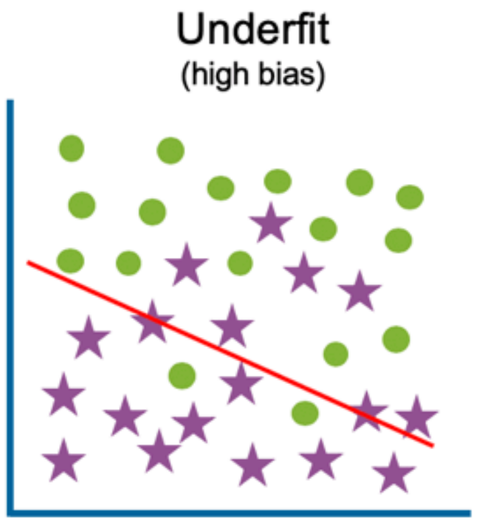
also known as independent variables or predictors, are the input variables used to predict or classify the target feature.These are the characteristics or attributes that provide information about the data.

Target, Descriptive Features?!

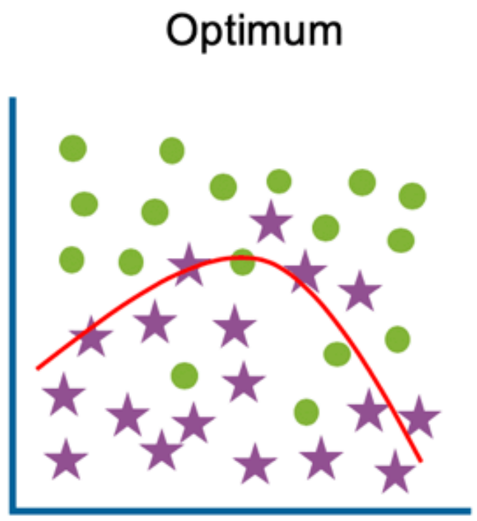




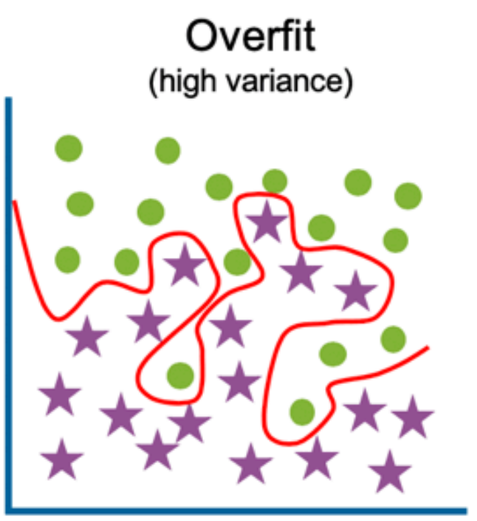
Overfiting & Underfiting



High training error
High test error

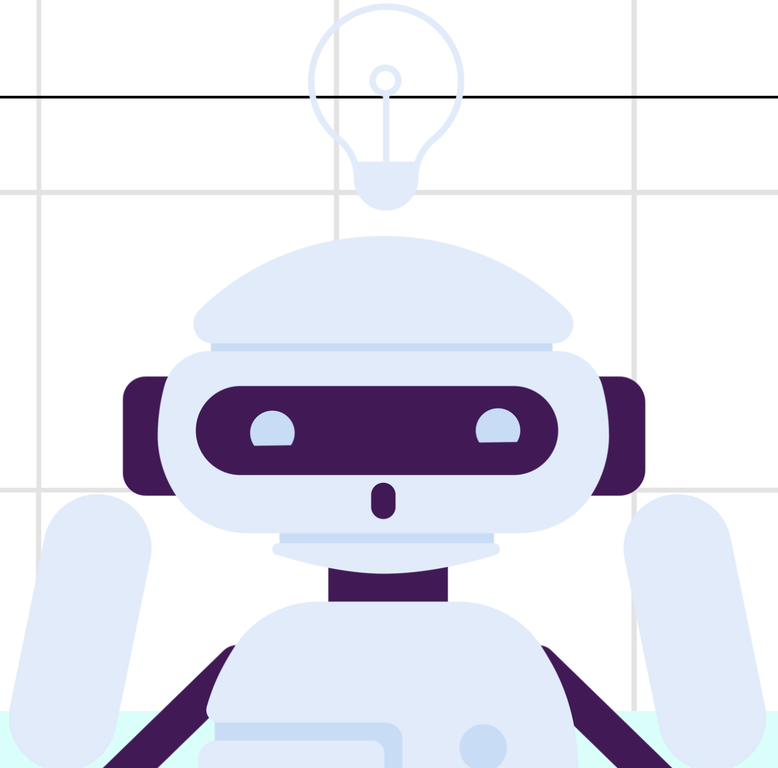


Low training error
Low test error



Low training error
High test error

**

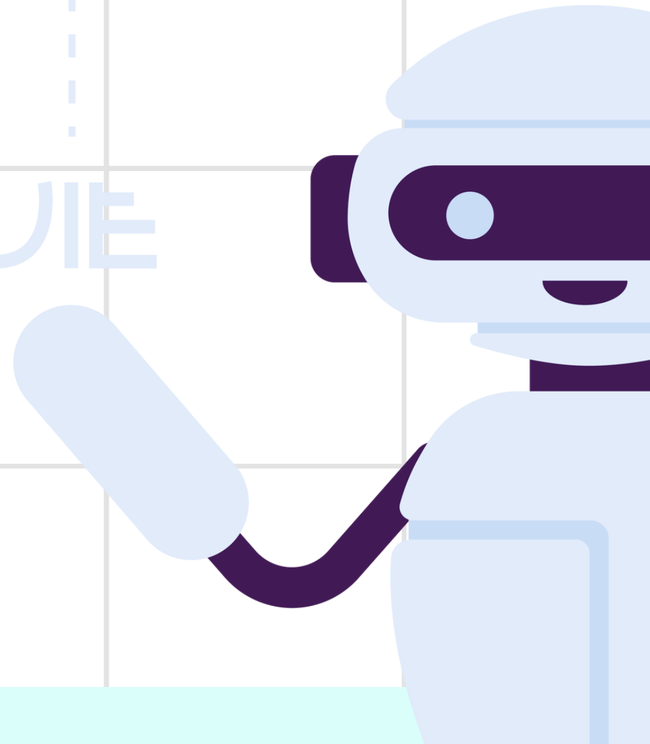


It’s Time to FOCUS

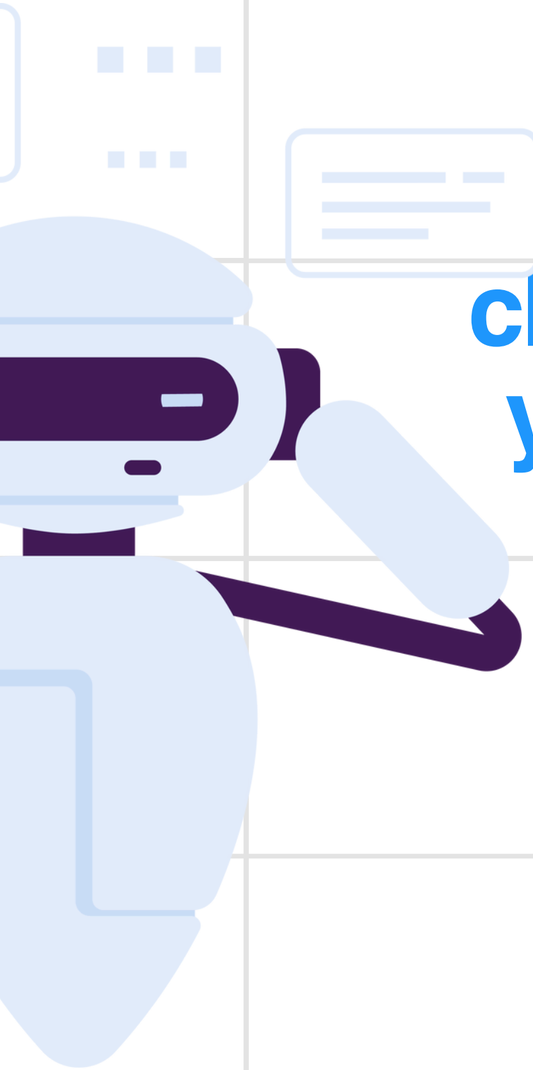
Sofia

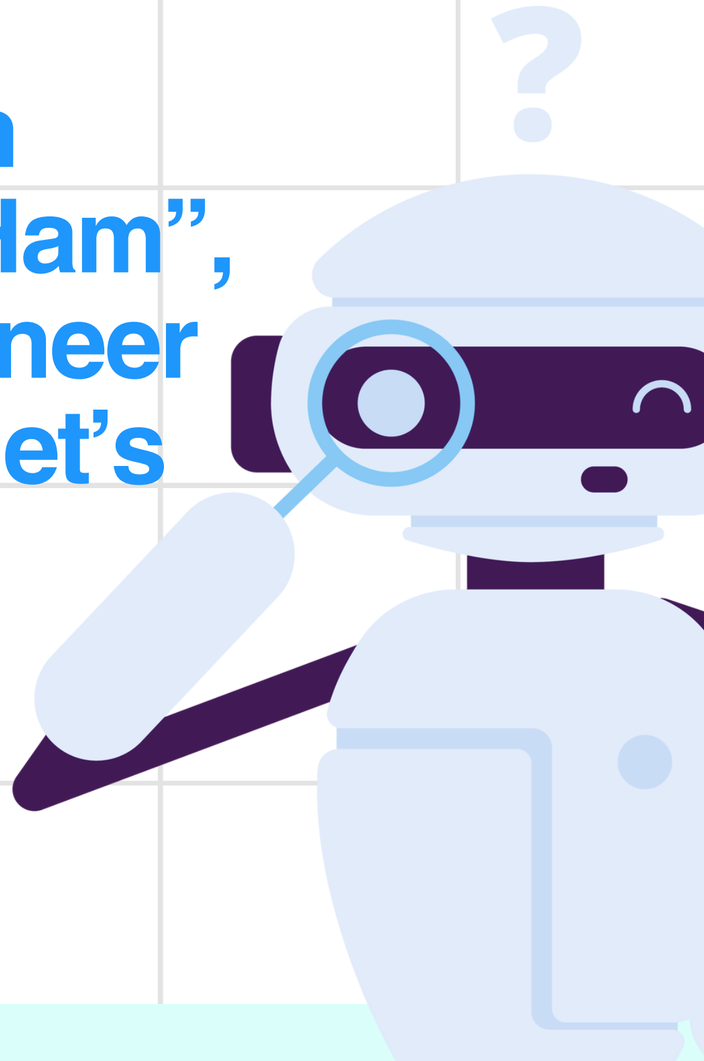
Has diabete

Healthy



The most used example in classification is “Spam” and “Ham”, you can’t say you are ML engineer without knowing that XD. So let’s follow the classical way :)

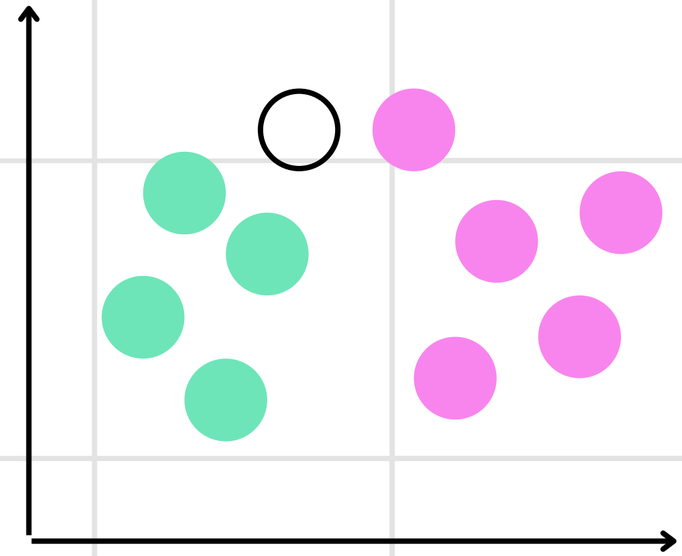


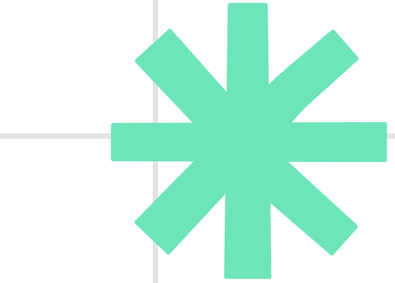


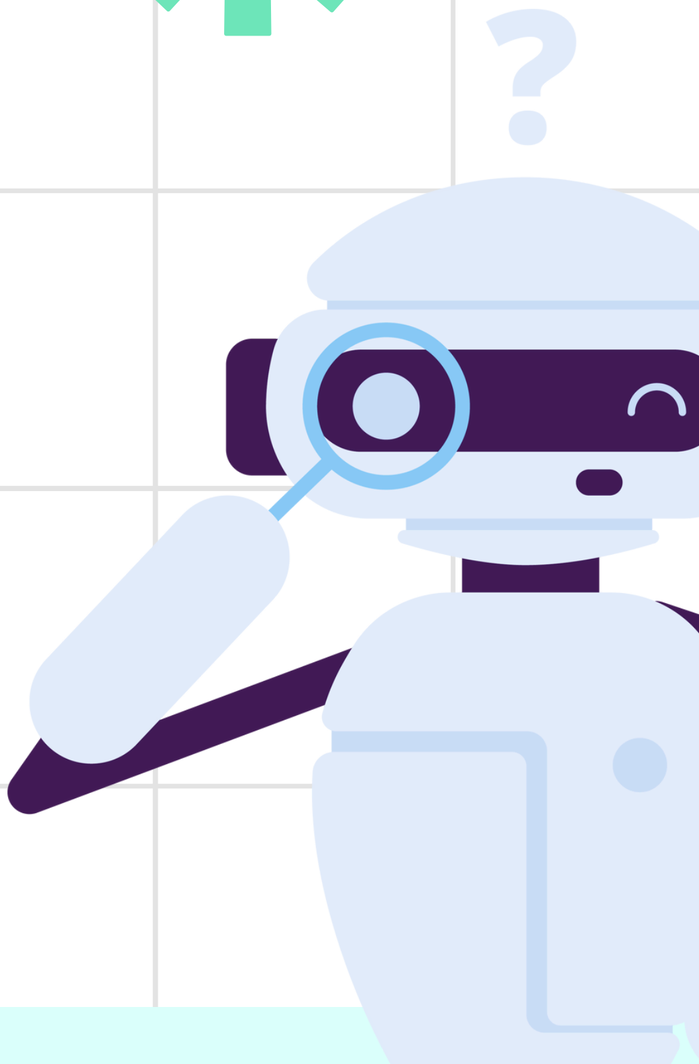
k-Nearest Neighbor k-NN

WTM

I’m just a girl 🎀





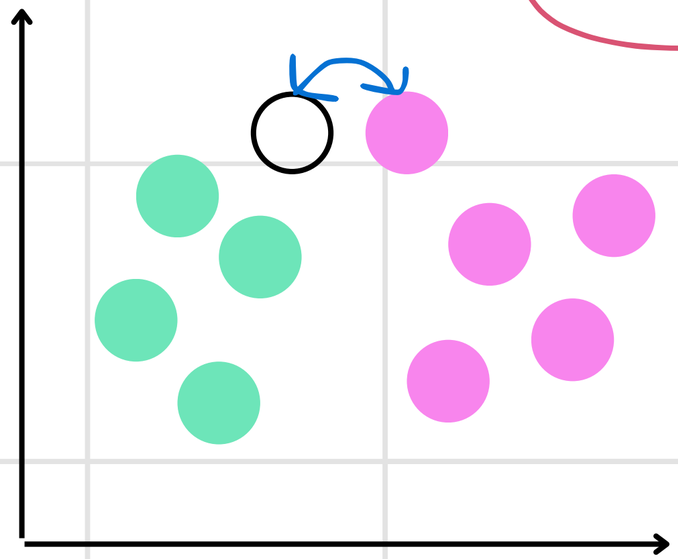


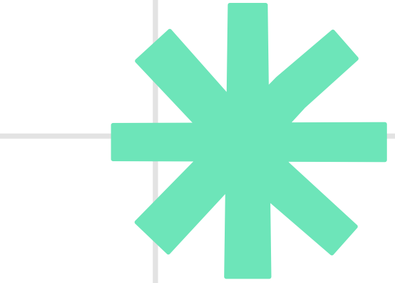
k-Nearest Neighbor k-NN

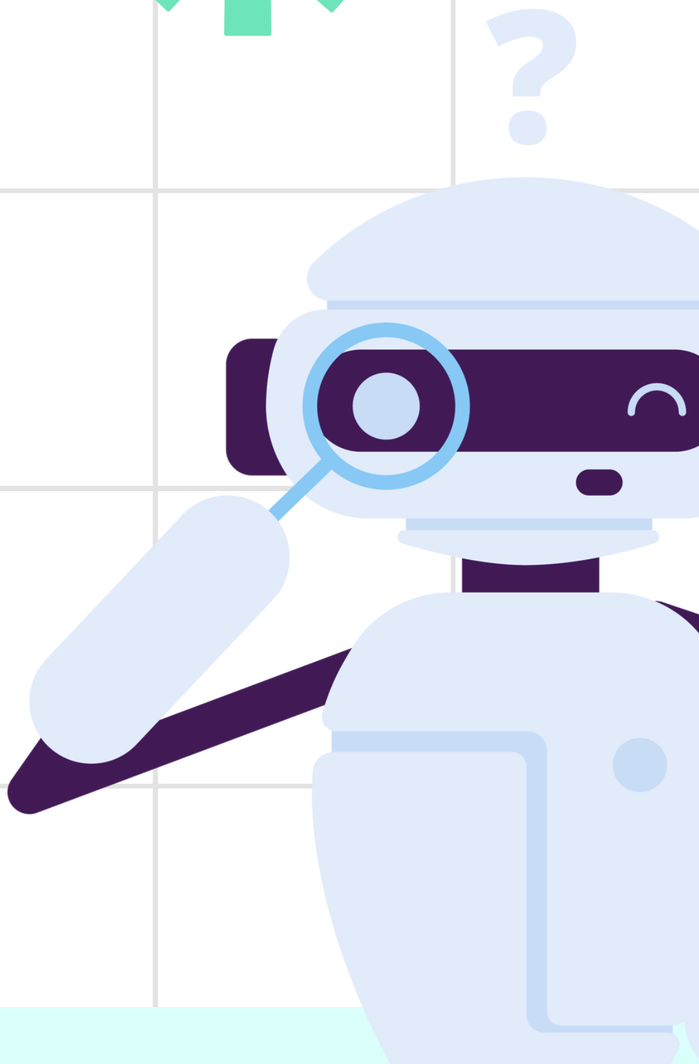
k = 1

WTM

I’m just a girl 🎀





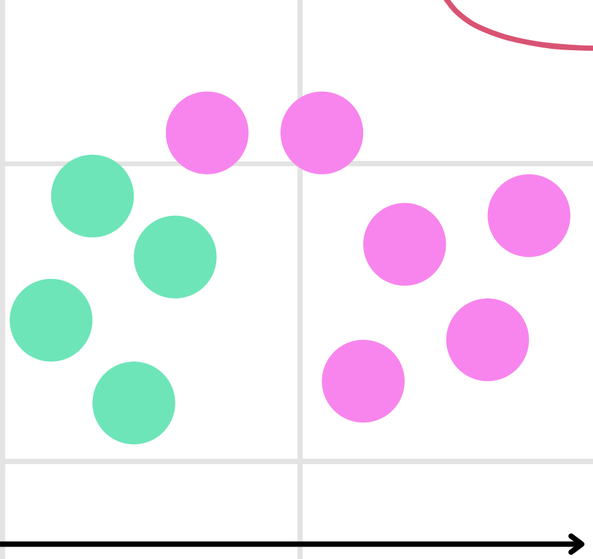


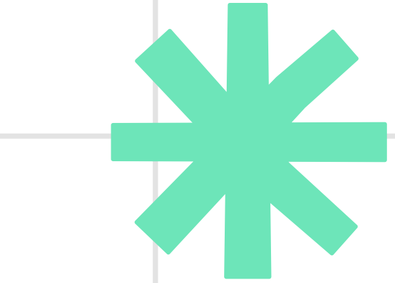
k-Nearest Neighbor k-NN

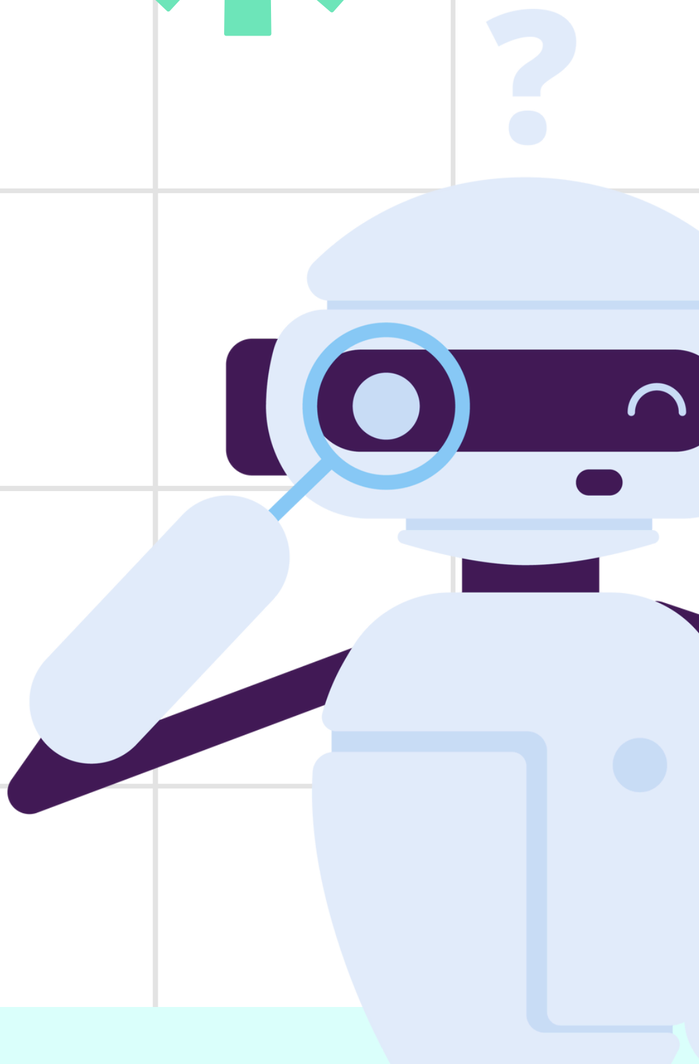
k = 1

WTM

I’m just a girl 🎀







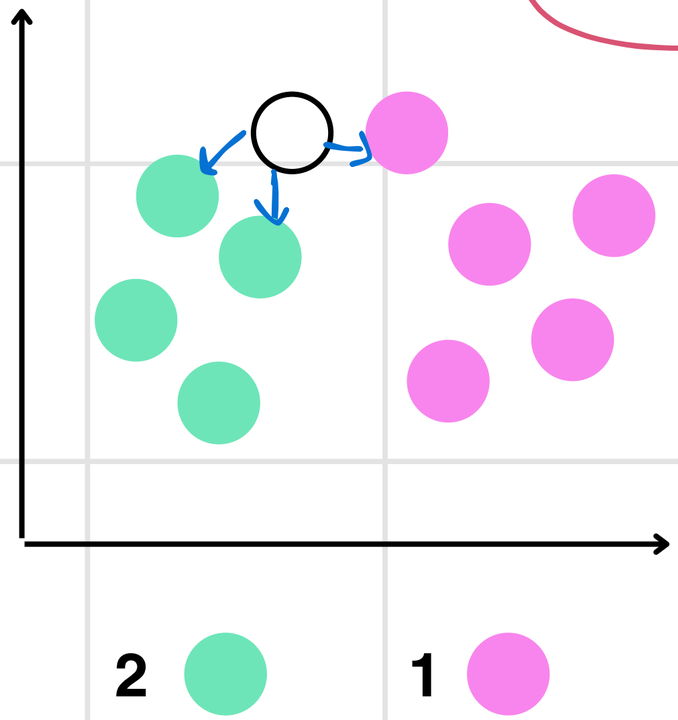
Our problem is finding our friend “k”

k-Nearest Neighbor k-NN

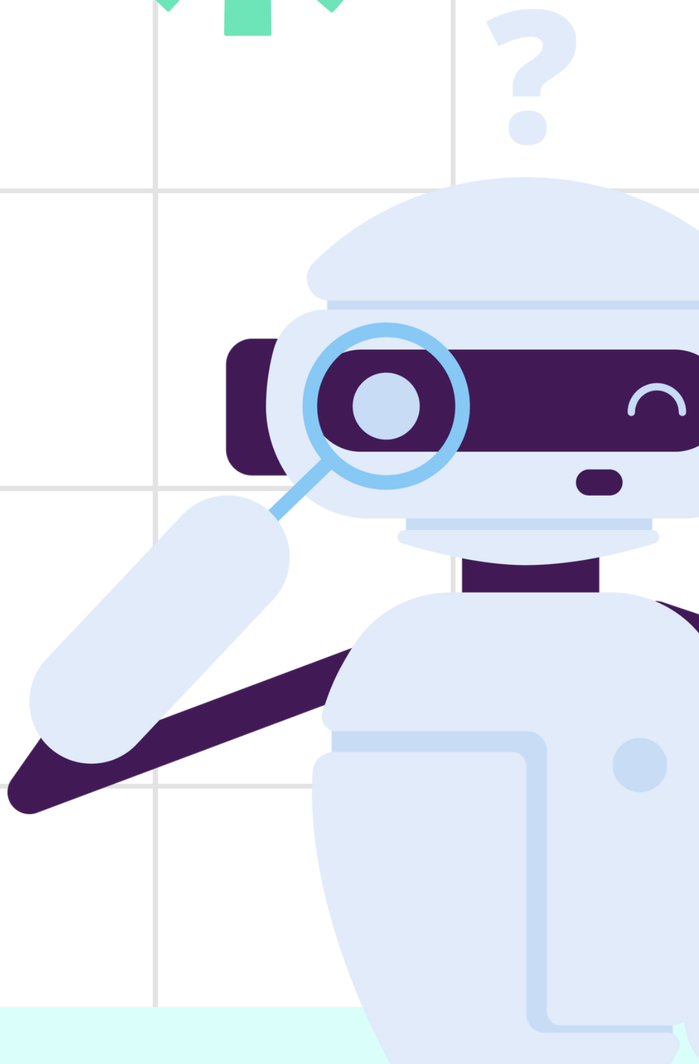
k = 3

WTM

I’m just a girl 🎀



2 1



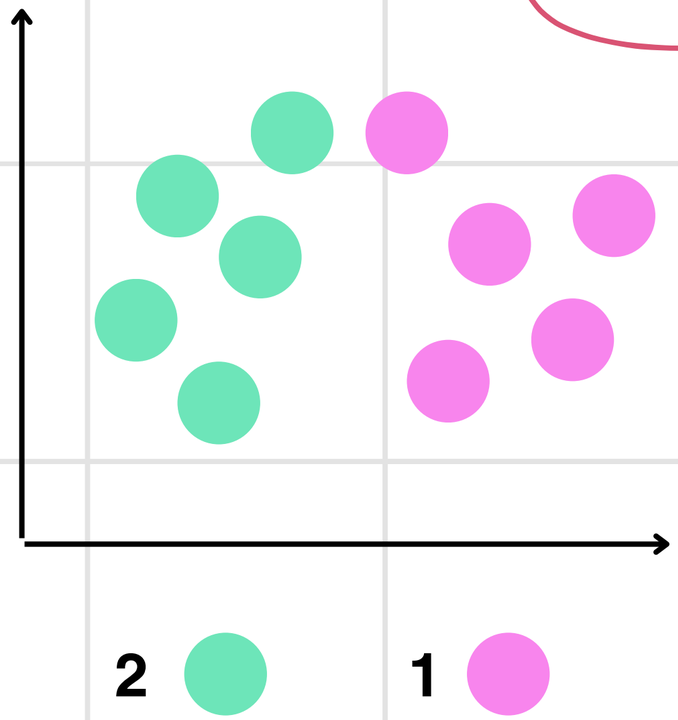
Our problem is finding our friend “k”

k-Nearest Neighbor k-NN

k = 3

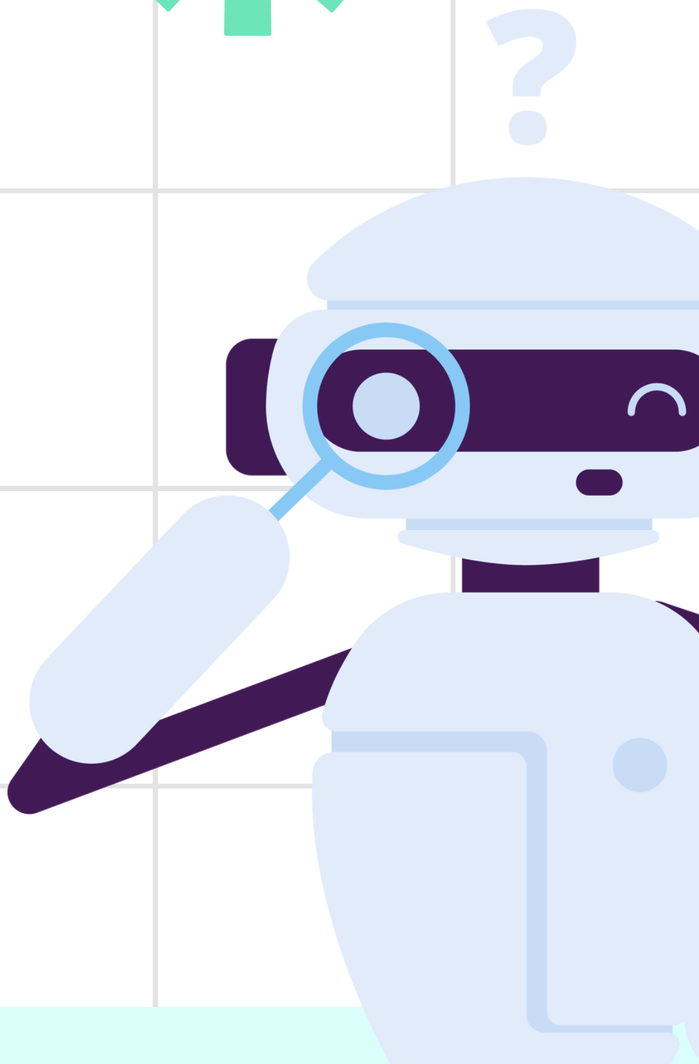
WTM

I’m just a girl 🎀



2 1

WTM wins 🎉

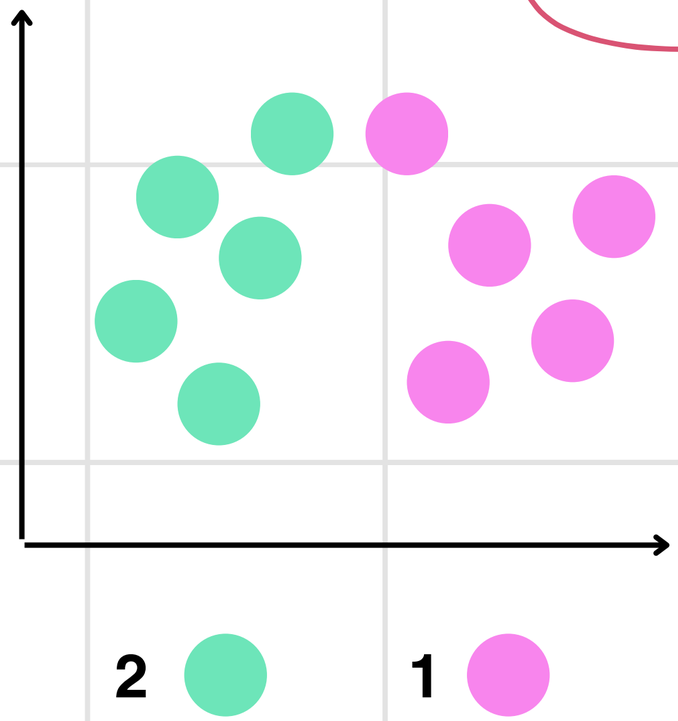


k-Nearest Neighbor k-NN

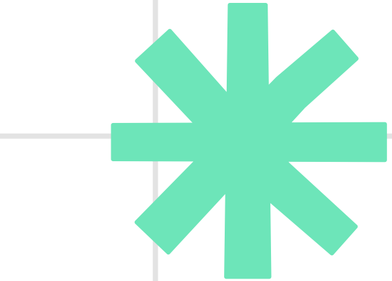
k = 3

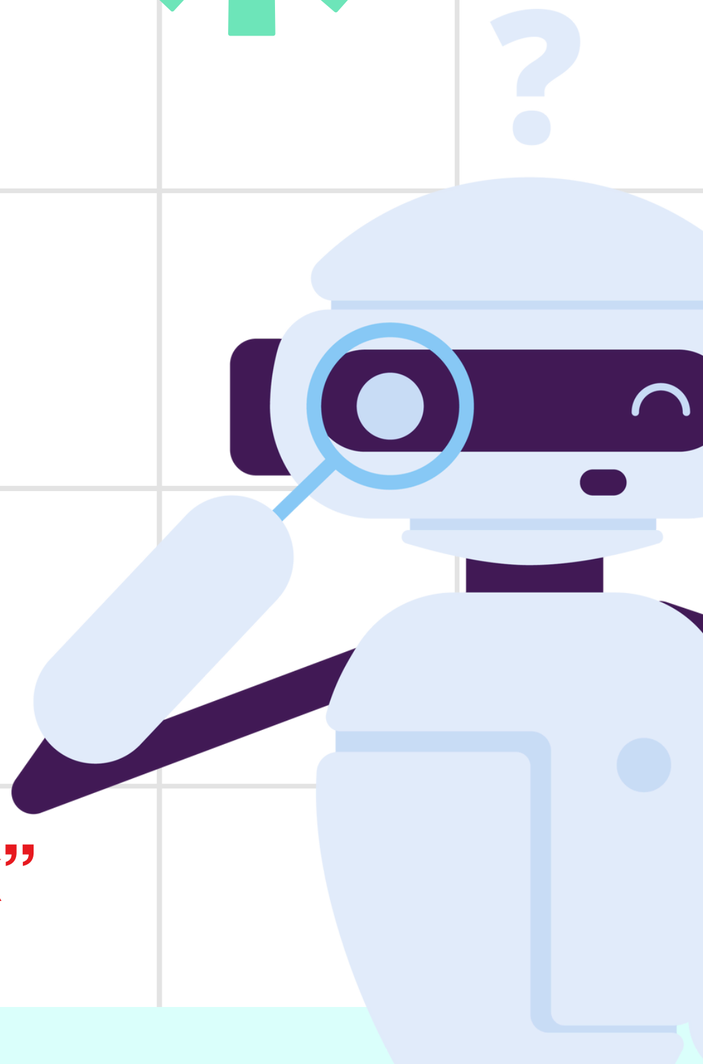
WTM

I’m just a girl 🎀



2 1





Soo.... Our mission is finding our friend “k”

How does k-NN work in practice?

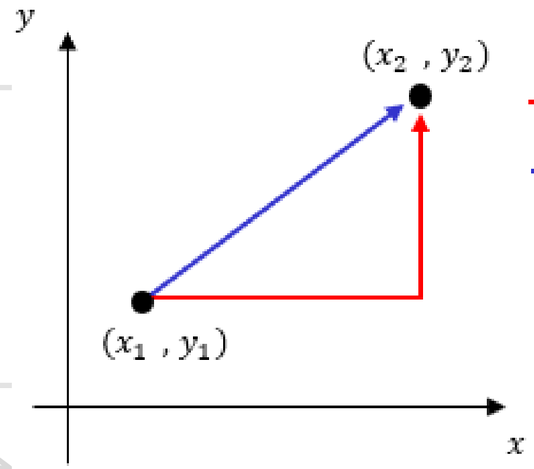
Similarity metrics

Euclidean Distance

$$\sqrt{\sum_{i=0}^{n}(x_i - y_i)^2}$$

Manhattan Distance

$$\sum_{i=1}^{n} |x_i - y_i|$$



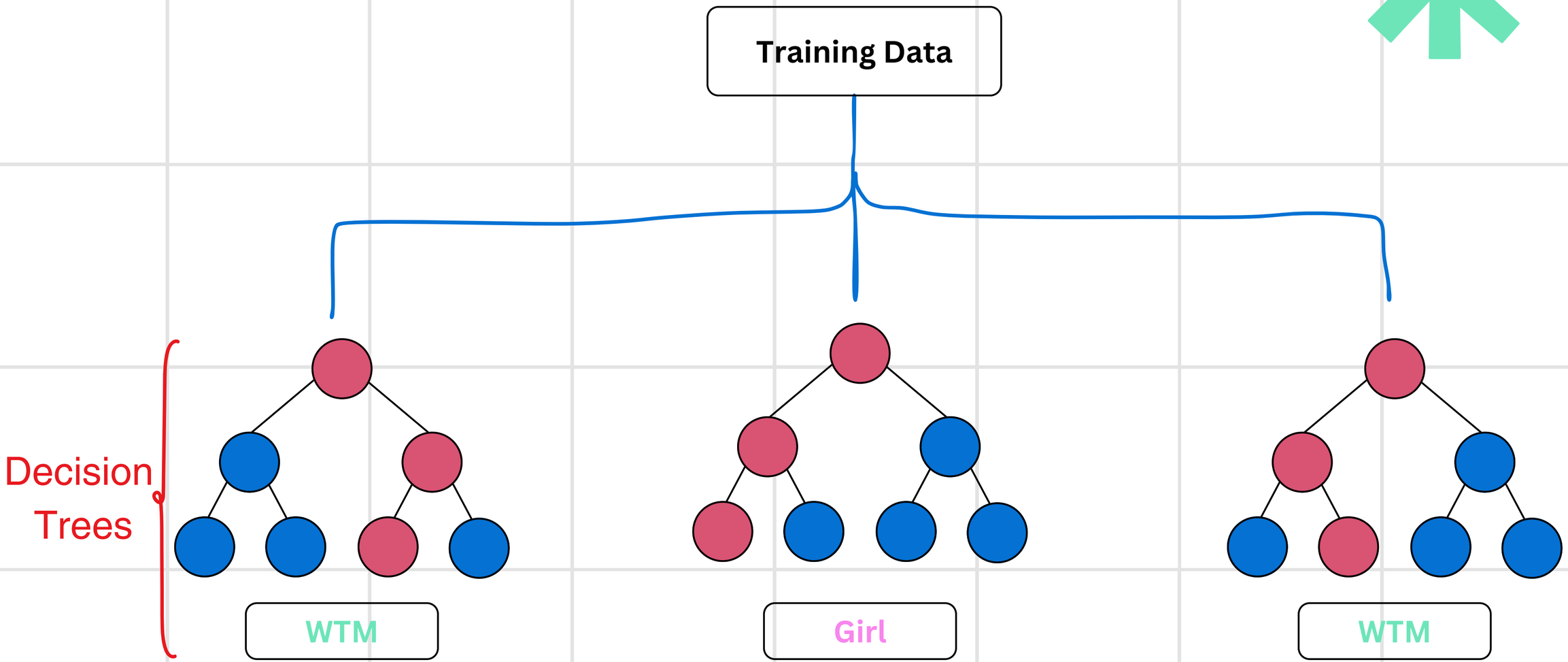
Manhattan Distance L^1

Euclidean Distance L^2

$$L^1 = |x_2 - x_1| + |y_2 - y_1|$$

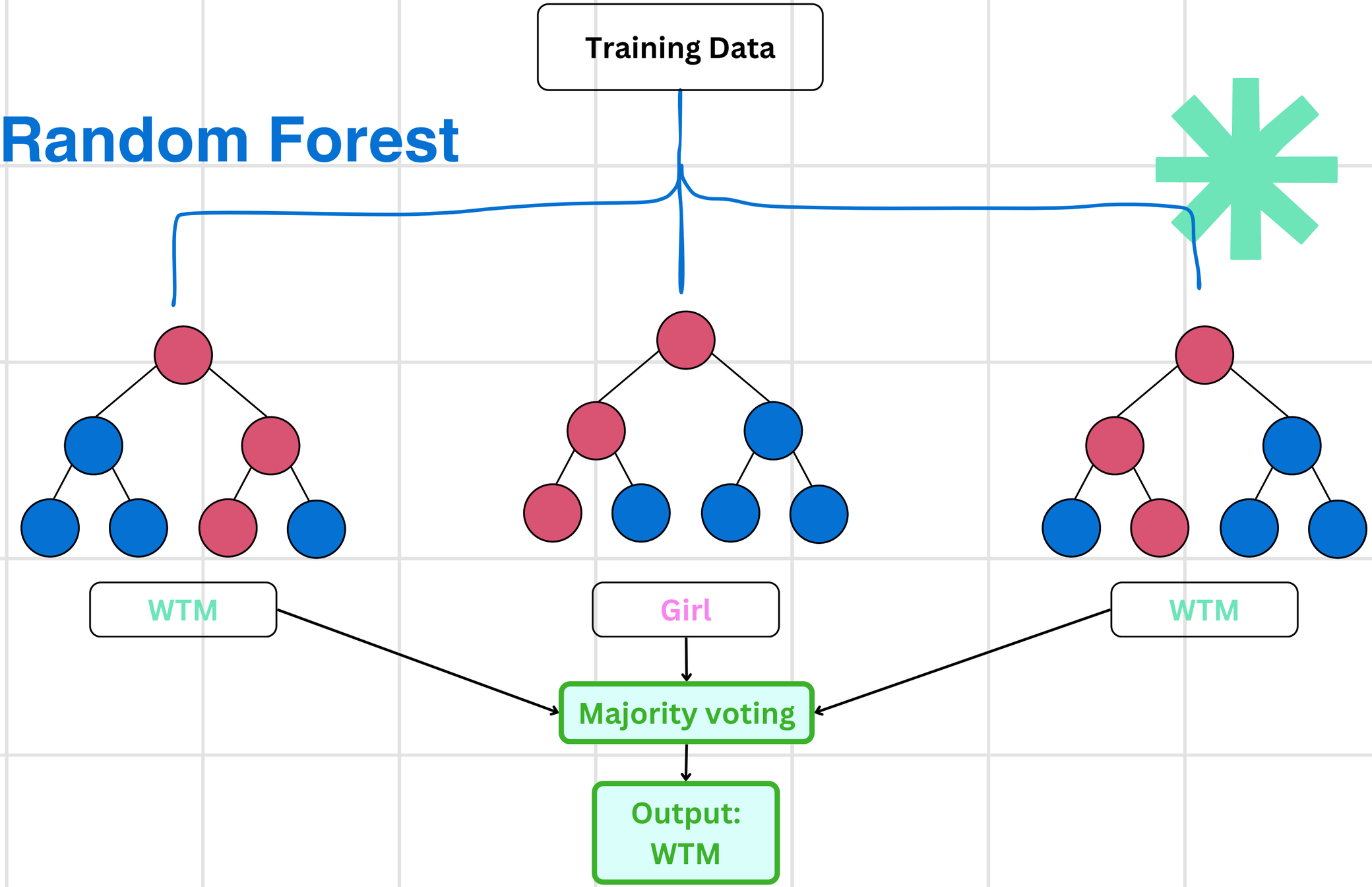
$$L^2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Random Forest

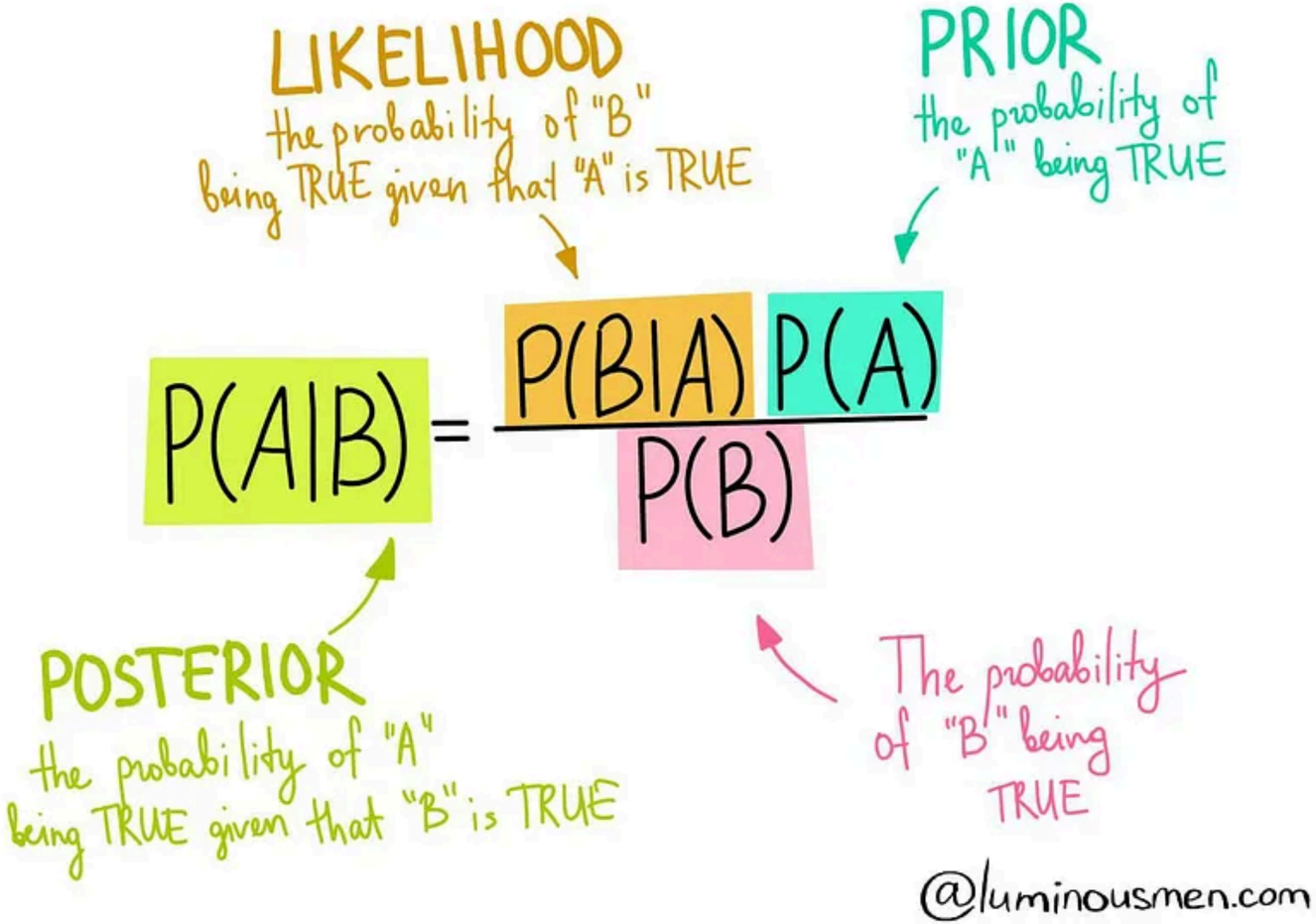


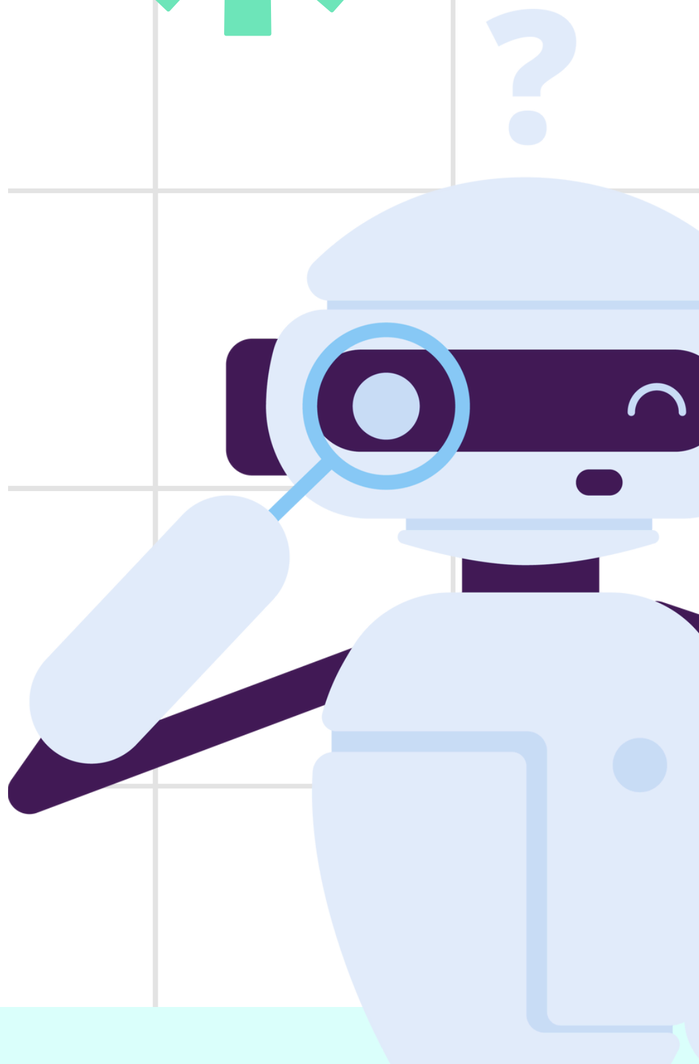
How to choose?

Random Forest



The Naive Bayes' Classifier





The Naive Bayes' Classifier

ID	Credit History	Guarantor/ CoApplicant	Accomodation	Fraud
1	current	none	own	true
2	paid	none	own	false
3	paid	none	own	false
4	paid	guarantor	rent	true
5	arrears	none	own	false
6	arrears	none	own	true
7	current	none	own	false
8	arrears	none	own	false
9	current	none	rent	false
10	none	none	own	true
11	current	coapplicant	own	false
12	current	none	own	true
13	current	none	rent	true
14	paid	none	own	false
15	arrears	none	own	false
16	current	none	own	false
17	arrears	coapplicant	rent	false
18	arrears	none	free	false
19	arrears	none	own	false
20	paid	none	own	false

The probabilities needed by a Naive Bayes prediction model

$P(fr)$	$=$	0.3	$P(\neg fr)$	$=$	0.7
$P(\text{CH} = \textit{'none'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \textit{'none'} \mid \neg fr)$	$=$	0
$P(\text{CH} = \textit{'paid'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \textit{'paid'} \mid \neg fr)$	$=$	0.2857
$P(\text{CH} = \textit{'current'} \mid fr)$	$=$	0.5	$P(\text{CH} = \textit{'current'} \mid \neg fr)$	$=$	0.2857
$P(\text{CH} = \textit{'arrears'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \textit{'arrears'} \mid \neg fr)$	$=$	0.4286
$P(\text{GC} = \textit{'none'} \mid fr)$	$=$	0.8334	$P(\text{GC} = \textit{'none'} \mid \neg fr)$	$=$	0.8571
$P(\text{GC} = \textit{'guarantor'} \mid fr)$	$=$	0.1666	$P(\text{GC} = \textit{'guarantor'} \mid \neg fr)$	$=$	0
$P(\text{GC} = \textit{'coapplicant'} \mid fr)$	$=$	0	$P(\text{GC} = \textit{'coapplicant'} \mid \neg fr)$	$=$	0.1429
$P(\text{ACC} = \textit{'own'} \mid fr)$	$=$	0.6666	$P(\text{ACC} = \textit{'own'} \mid \neg fr)$	$=$	0.7857
$P(\text{ACC} = \textit{'rent'} \mid fr)$	$=$	0.3333	$P(\text{ACC} = \textit{'rent'} \mid \neg fr)$	$=$	0.1429
$P(\text{ACC} = \textit{'free'} \mid fr)$	$=$	0	$P(\text{ACC} = \textit{'free'} \mid \neg fr)$	$=$	0.0714

The Naive Bayes' Classifier

CREDIT HISTORY	GUARANTOR/COAPPLICANT	ACCOMODATION	FRAUDULENT
paid	none	rent	?

$$\left(\prod_{k=1}^{m} P(\mathbf{q}[k] \mid fr)\right) \times P(fr) = 0.0139$$

$$\left(\prod_{k=1}^{m} P(\mathbf{q}[k] \mid \neg fr)\right) \times P(\neg fr) = 0.0245$$

The probabilities needed by a Naive Bayes prediction model

$P(fr)$	$=$	0.3	$P(\neg fr)$	$=$	0.7
$P(\text{CH} = \text{'none'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \text{'none'} \mid \neg fr)$	$=$	0
$P(\text{CH} = \text{'paid'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \text{'paid'} \mid \neg fr)$	$=$	0.2857
$P(\text{CH} = \text{'current'} \mid fr)$	$=$	0.5	$P(\text{CH} = \text{'current'} \mid \neg fr)$	$=$	0.2857
$P(\text{CH} = \text{'arrears'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \text{'arrears'} \mid \neg fr)$	$=$	0.4286
$P(\text{GC} = \text{'none'} \mid fr)$	$=$	0.8334	$P(\text{GC} = \text{'none'} \mid \neg fr)$	$=$	0.8571
$P(\text{GC} = \text{'guarantor'} \mid fr)$	$=$	0.1666	$P(\text{GC} = \text{'guarantor'} \mid \neg fr)$	$=$	0
$P(\text{GC} = \text{'coapplicant'} \mid fr)$	$=$	0	$P(\text{GC} = \text{'coapplicant'} \mid \neg fr)$	$=$	0.1429
$P(\text{ACC} = \text{'own'} \mid fr)$	$=$	0.6666	$P(\text{ACC} = \text{'own'} \mid \neg fr)$	$=$	0.7857
$P(\text{ACC} = \text{'rent'} \mid fr)$	$=$	0.3333	$P(\text{ACC} = \text{'rent'} \mid \neg fr)$	$=$	0.1429
$P(\text{ACC} = \text{'free'} \mid fr)$	$=$	0	$P(\text{ACC} = \text{'free'} \mid \neg fr)$	$=$	0.0714

The Naive Bayes' Classifier

CREDIT HISTORY	GUARANTOR/COAPPLICANT	ACCOMODATION	FRAUDULENT
paid	none	rent	?

$$\left(\prod_{k=1}^{m} P(\mathbf{q}[k] \mid fr)\right) \times P(fr) = 0.0139$$

$$\left(\prod_{k=1}^{m} P(\mathbf{q}[k] \mid \neg fr)\right) \times P(\neg fr) = 0.0245$$

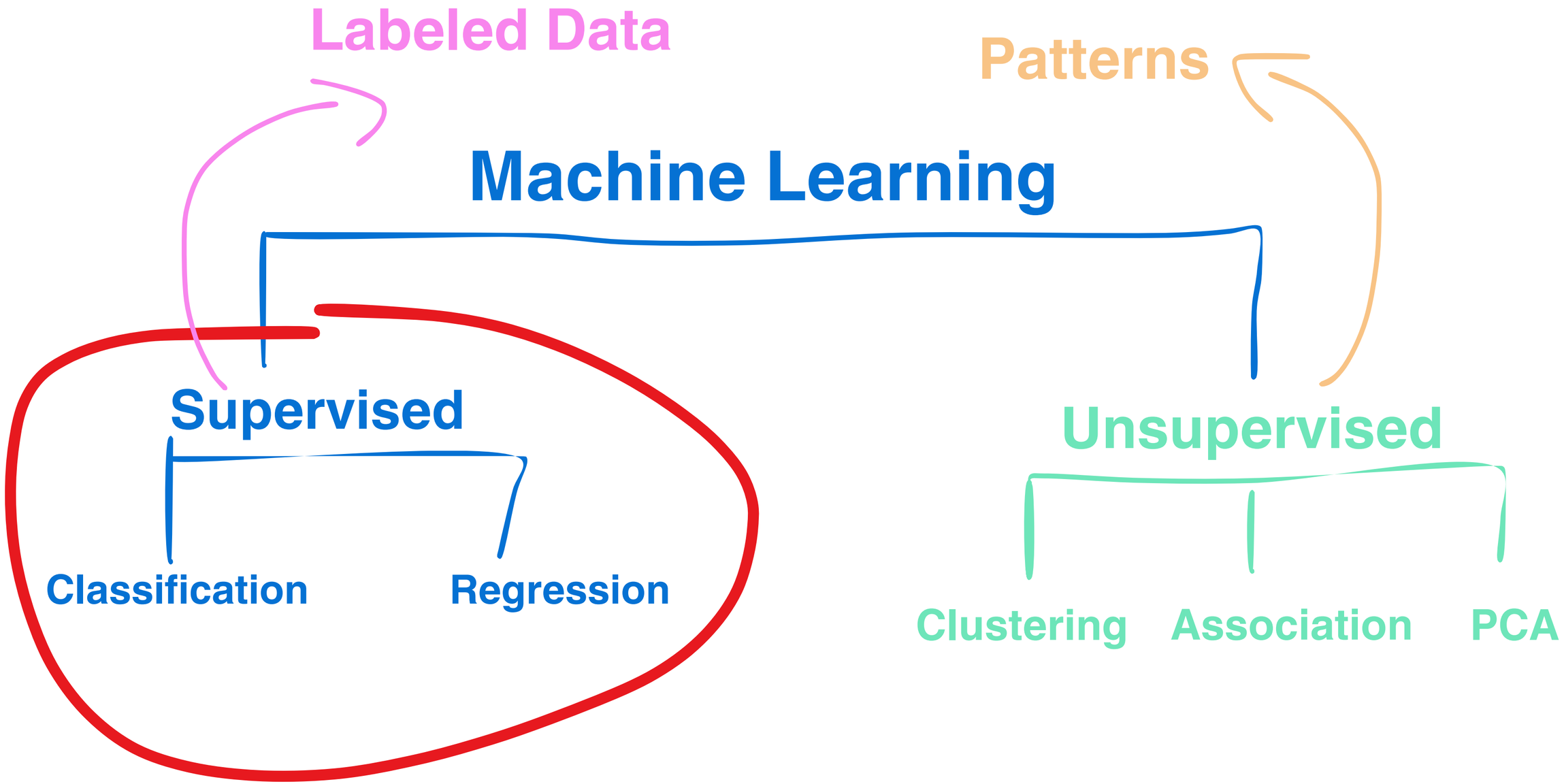
false

The probabilities needed by a Naive Bayes prediction model

$P(fr)$	=	0.3	$P(\neg fr)$	=	0.7
$P(\text{CH} = \text{'none'} \mid fr)$	=	0.1666	$P(\text{CH} = \text{'none'} \mid \neg fr)$	=	0
$P(\text{CH} = \text{'paid'} \mid fr)$	=	0.1666	$P(\text{CH} = \text{'paid'} \mid \neg fr)$	=	0.2857
$P(\text{CH} = \text{'current'} \mid fr)$	=	0.5	$P(\text{CH} = \text{'current'} \mid \neg fr)$	=	0.2857
$P(\text{CH} = \text{'arrears'} \mid fr)$	=	0.1666	$P(\text{CH} = \text{'arrears'} \mid \neg fr)$	=	0.4286
$P(\text{GC} = \text{'none'} \mid fr)$	=	0.8334	$P(\text{GC} = \text{'none'} \mid \neg fr)$	=	0.8571
$P(\text{GC} = \text{'guarantor'} \mid fr)$	=	0.1666	$P(\text{GC} = \text{'guarantor'} \mid \neg fr)$	=	0
$P(\text{GC} = \text{'coapplicant'} \mid fr)$	=	0	$P(\text{GC} = \text{'coapplicant'} \mid \neg fr)$	=	0.1429
$P(\text{ACC} = \text{'own'} \mid fr)$	=	0.6666	$P(\text{ACC} = \text{'own'} \mid \neg fr)$	=	0.7857
$P(\text{ACC} = \text{'rent'} \mid fr)$	=	0.3333	$P(\text{ACC} = \text{'rent'} \mid \neg fr)$	=	0.1429
$P(\text{ACC} = \text{'free'} \mid fr)$	=	0	$P(\text{ACC} = \text{'free'} \mid \neg fr)$	=	0.0714

Thank you!

See you in the next session :)


Labeled Data
Patterns
Machine Learning
Supervised
Unsupervised
Classification
Regression
Clustering
Association
PCA