Mathematics for Computer Science
CSE 401
Naïve Bayes Method/Classifier/Theorem/Algorithm
Data Analytics

Shah Murtaza Rashid Al Masud, PhD
Department of CSE
University of Asia Pacific

Naïve Bayes' Classification Algorithm

Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems.

Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.

It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

Naïve Bayes' Classification Algorithm

Why is it called Naïve Bayes?

The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as:

Naïve: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.

Bayes: It is called Bayes because it depends on the principle of Bayes' Theorem

Naïve Bayes' Classification Algorithm

Bayes' theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

First, let's take a formula of conditional probability, and try to derive Bayes Theorem:

$$P(B|A) = P(A \cap B)/P(B),$$

where probability of B given A, meaning what is the probability of event B when event A is already taken place, equals probability of A intersection B (meaning the probability of both A and B events taken place) divided by probability of B.

Naïve Bayes' Classification Algorithm

or $P(A|B) = P(B \cap A)/P(A)$,

where probability of A given B, meaning what is the probability of event A when event B is already taken place, equals probability of B intersection A (meaning the probability of both B and A events taken place) divided by probability of A.

Let's take a closer look, we see that $P(A \cap B)$ and $P(B \cap A)$ are basically the same, so we can write them as $P(A \cap B) = P(B \cap A)$. Since they are the same, we can get two formulas and move denominator to the left of the equation:

$P(A \cap B) = P(A|B) * P(B)$, and $P(B \cap A) = P(B|A) * P(A)$ and equate them:

$P(A|B) * P(B) = P(B|A) * P(A)$.

Naïve Bayes' Classification Algorithm

So, when we want to find probability of A given B we can write our equation this way:

P(A|B) = P(B|A) * P(A) / P(B), and this is the equation of Bayes Theorem.

Applying Bayes Theorem Equation in Algorithm

Let's break down our equation and understand how it works:

Probability of A given B

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

P(A|B) is Posterior probability (the probability of A given B): Probability of hypothesis A on the observed event B or is the probability of the hypothesis given that the evidence is there

P(B|A) is Likelihood probability (the probability of B given A): Probability of the evidence given that the probability of a hypothesis is true

P(A) is Prior Probability: Probability of hypothesis before observing the evidence or is the probability of hypothesis H being true

P(B) is Marginal Probability: Probability of Evidence (regardless of the hypothesis)

Naïve Bayes’ Classification Algorithm

What is Hypothesis?

Hypothesis is an assumption that is made on the basis of some evidence. This is the initial point of any investigation that translates the research questions into a prediction. It includes components like variables, population and the relation between the variables. A research hypothesis is a hypothesis that is used to test the relationship between two or more variables.

Naïve Bayes' Classification Algorithm

Likelihood of the Evidence given that the Hypothesis is True

Prior Probability of the Hypothesis

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Posterior Probability of the Hypothesis given that the Evidence is True

Prior Probability that the evidence is True

For many predictors, we can formulate the posterior probability as follows:

P(A|B) = P(B1|A) * P(B2|A) * P(B3|A) * P(B4|A) …

Naïve Bayes' Classification Algorithm

As we know $$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$ $$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

P(A|B) is Posterior probability (the probability of A given B): Probability of hypothesis A on the observed event B or is the probability of the hypothesis given that the evidence is there (the conditional probability of the response variable that belongs to a particular value, given the input attributes)

P(B|A) is Likelihood probability (the probability of B given A): Probability of the evidence given that the probability of a hypothesis is true (this is termed as the likelihood of the training data)

P(A) is Prior Probability: Probability of hypothesis before observing the evidence or is the probability of hypothesis H being true (the prior probability of the response variable)

P(B) is Marginal Probability: Probability of Evidence (regardless of the hypothesis) or the probability of training data (input attributes)

The Bayes' Theorem can be reformulated in correspondence with the machine learning algorithm as:

```
posterior = (prior x likelihood) / (evidence)
```

Naïve Bayes' Classification Algorithm

The likelihood can be written considering n different attributes as:

$$P(X_1 \ldots X_n | Y) = \prod_{i=1}^{n} P(X_i | Y)$$

Bayes rule is a way to find P(Y|X) from P(X|Y)

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} \qquad P(Y|X) = \frac{P(X \cap Y)}{P(X)}$$

P(Evidence|Outcome)
Knowing from Training Data

$\rightarrow$

P(Outcome|Evidence)
To be predicted for test data

Naïve Bayes' Classification Algorithm

If **X** represents **n** different parameters/features then

$$X = (x_1, x_2, x_3, \ldots\ldots, x_n)$$

By expanding using the chain rule we get,

$$P(y|x_1, \ldots, x_n) = \frac{P(x_1|y)P(x_2|y)\ldots P(x_n|y)P(y)}{P(x_1)P(x_2)\ldots P(x_n)}$$

which can be expressed as:

$$P(y|x_1, \ldots, x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1)P(x_2)\ldots P(x_n)}$$

Now, as the denominator remains constant for a given input, we can remove that term:

$$P(y|x_1, \ldots, x_n) \propto P(y)\prod_{i=1}^{n} P(x_i|y)$$

Naïve Bayes' Classification Algorithm

In order to calculate posterior probability, first we need to

(i) calculate frequency for each attribute against the target. Then,

(ii) transform the frequency to likelihood values and

(iii) use the Naive Bayesian equation to calculate the posterior probability for each class.

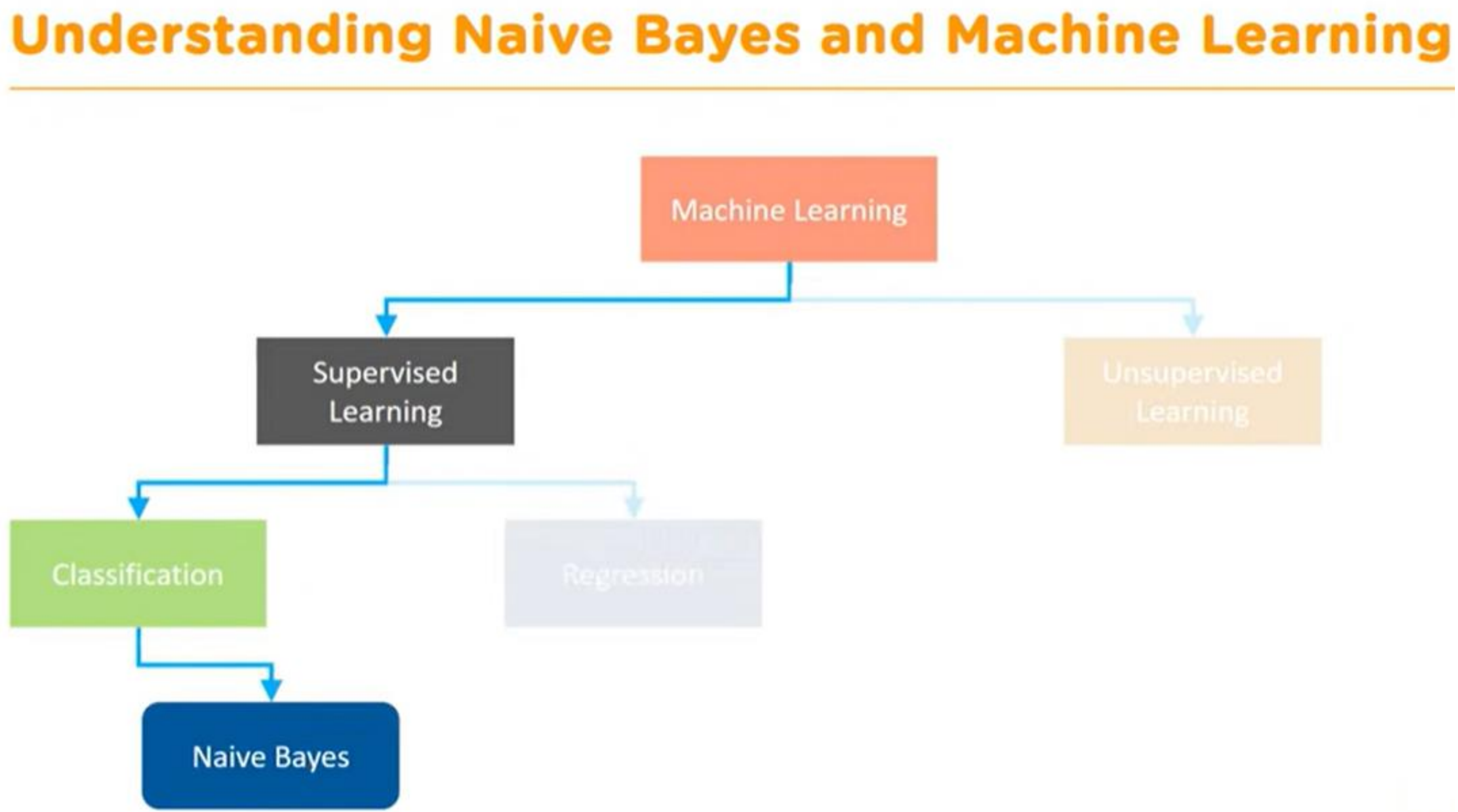
(iv) Perform normalization

(v) The class with the highest posterior probability is the outcome of prediction.

Naïve Bayes' Classification Algorithm

BAYES' THEOREM BASICALLY CALCULATES THE CONDITIONAL PROBABILITY OF THE OCCURRENCE OF AN EVENT BASED ON PRIOR KNOWLEDGE OF CONDITIONS THAT MIGHT BE RELATED TO THE EVENT

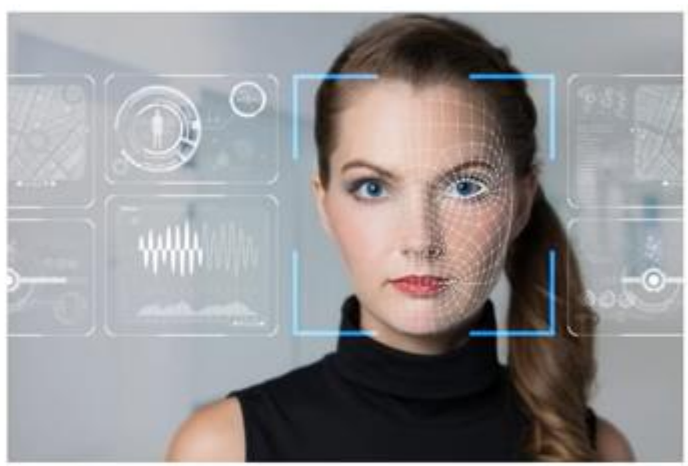
Naïve Bayes’ Classification Algorithm



Naïve Bayes' Classification Algorithm

Where is Naive Bayes used?

Face Recognition



Weather Prediction

Naïve Bayes' Classification Algorithm

Where is Naive Bayes used?

Medical Diagnosis

News Classification

Naïve Bayes' Classification Algorithm

Naive Bayes can be used for various things like face recognition, weather prediction, Medical Diagnosis, News classification, Sentiment Analysis, and a lot more.

Naïve Bayes' Classification Algorithm

Understanding Naive Bayes Classifier

Naive Bayes Classifier is based on Bayes' Theorem which gives the conditional probability of an event A given B

Bayes Theorem

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

where:

$P(A|B)$ = Conditional Probability of A given B

$P(B|A)$ = Conditional Probability of A given B

$P(A)$ = Probability of event A

$P(B)$ = Probability of event A

Naïve Bayes' Classification Algorithm

Naïve Bayes Example: Let's take a dataset to predict whether we can *pet an animal or not.*

	Animals	Size of Animal	Body Color	Can we Pet them
0	Dog	Medium	Black	Yes
1	Dog	Big	White	No
2	Rat	Small	White	Yes
3	Cow	Big	White	Yes
4	Cow	Small	Brown	No
5	Cow	Big	Black	Yes
6	Rat	Big	Brown	No
7	Dog	Small	Brown	Yes
8	Dog	Medium	Brown	Yes
9	Cow	Medium	White	No
10	Dog	Small	Black	Yes
11	Rat	Medium	Black	No
12	Rat	Small	Brown	No
13	Cow	Big	White	Yes

Naïve Bayes' Classification Algorithm

Naïve Bayes Example: Let's take a dataset to predict whether we can *pet an animal or not.*

Now if we send our test data, suppose **test = (Cow, Medium, Black)**

Probability of petting an animal :

$$P(Yes|Test) = \frac{P(Animal = Cow|Yes) * P(Size = Medium|Yes) * P(Color = Black|Yes) * P(Yes)}{P(Test)}$$

And the probability of not petting an animal:

$$P(No|Test) = \frac{P(Animal = Cow|No) * P(Size = Medium|No) * P(Color = Black|No) * P(No)}{P(Test)}$$

Naïve Bayes' Classification Algorithm

Frequency and likelihood tables of the dataset:

Animals

	Yes	No	P(Yes)	P(No)
Dog	4	1	4/8	1/6
Rat	1	3	1/8	3/6
Cow	3	2	3/8	2/6
Total	8	6	100%	100%

Size of Animal

	Yes	No	P(Yes)	P(No)
Medium	2	2	2/8	2/6
Big	3	2	3/8	2/6
Small	3	2	3/8	2/6
Total	8	6	100%	100%

Body Color

	Yes	No	P(Yes)	P(No)
Black	3	1	3/8	1/6
White	3	2	3/8	2/6
Brown	2	3	2/8	3/6
Total	8	6	100%	100%

We need to find $P(x_i|y_j)$ for each x_i in X and each y_j in Y. All these calculations have been demonstrated here

Naïve Bayes' Classification Algorithm

We also need the probabilities (P(y)), which are calculated in the table below. For example, P(Pet Animal = NO) = 6/14.

Play		P(yes)/P(no)
Yes	8	8/14
No	6	6/14
Total	14	100%

Naïve Bayes' Classification Algorithm

Now if we send our test data, suppose **test = (Cow, Medium, Black)**

Probability of petting an animal :

$$P(Yes|Test) = \frac{P(Animal = Cow|Yes) * P(Size = Medium|Yes) * P(Color = Black|Yes) * P(Yes)}{P(Test)}$$

$$P(YesITest) \ \alpha \ \frac{3}{8} * \frac{2}{8} * \frac{3}{8} * \frac{8}{14} \ \approx 0.0200$$

And the probability of not petting an animal:

$$P(No|Test) = \frac{P(Animal = Cow|No) * P(Size = Medium|No) * P(Color = Black|No) * P(No)}{P(Test)}$$

$$P(NoITest) \ \alpha \ \frac{2}{6} * \frac{2}{6} * \frac{1}{6} * \frac{6}{14} \ \approx 0.0079$$

Naïve Bayes' Classification Algorithm

No since, P(YesITest) + P(NoITest) = 1

These numbers can be converted into a probability by making the sum equal to 1 (normalization):

$$P(\text{YesITest}) = \frac{0.0200}{0.0200+0.0079} = 0.7168=0.72=72\%$$

$$P(\text{NoITest}) = \frac{0.0079}{0.0200+0.0079} = 0.2831=0.28=28\%$$

We see here that P(Yes|Test) > P(No|Test), so the prediction that we can pet this animal is **"Yes"**.

Naïve Bayes' Classification Algorithm

Working of Naïve Bayes' Classifier

Working of Naïve Bayes' Classifier can be understood with the help of the below example:

Suppose we have a dataset of **weather conditions** and corresponding target variable "**Play**". So using this dataset we need to decide that whether we should play or not on a particular day according to the weather conditions. So to solve this problem, we need to follow the below steps:

- Convert the given dataset into frequency tables.
- Generate Likelihood table by finding the probabilities of given features.
- Now, use Bayes theorem to calculate the posterior probability
- Perform normalization.

Problem: If the weather is sunny, then the Player should play or not?

Naïve Bayes' Classification Algorithm

Working of Naïve Bayes' Classifier

Solution: To solve this, first consider the below dataset:

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

Naïve Bayes' Classification Algorithm

Applying Bayes' theorem:

P(Yes|Sunny)= P(Sunny|Yes)*P(Yes)/P(Sunny)

P(Sunny|Yes)= 3/9= 0.33

P(Sunny)= 5/14 = 0.36

P(Yes)= 9/14= 0.64

So P(Yes|Sunny) = 0.33*0.64/0.36= **0.60**

P(No|Sunny)= P(Sunny|No)*P(No)/P(Sunny)

P(Sunny|NO)= 2/5=0.40

P(No)= 5/14 = 0.36

P(Sunny)= 5/14= 0.36

So P(No|Sunny)= 0.40*0.36/0.36 = **0.40**

So as we can see from the above calculation

that **P(Yes|Sunny)>P(No|Sunny)**

Hence on a Sunny day, Player can play the game.

Naïve Bayes' Classification Algorithm

Working of Naïve Bayes' Classifier

Let us take an example to get some better intuition. Consider the car theft problem with attributes Color, Type, Origin, and the target, Stolen can be either Yes or No.

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Naïve Bayes' Classification Algorithm

Working of Naïve Bayes' Classifier

Frequency and Likelihood tables of 'Color'

Frequency Table

		Stolen?	
		Yes	No
Color	Red	3	2
	Yellow	2	3

Likelihood Table

		Stolen?	
		P(Yes)	P(No)
Color	Red	3/5	2/5
	Yellow	2/5	3/5

Frequency and Likelihood tables of 'Type'

Frequency Table

		Stolen?	
		Yes	No
Type	Sports	4	2
	SUV	1	3

Likelihood Table

		Stolen?	
		P(Yes)	P(No)
Type	Sports	4/5	2/5
	SUV	1/5	3/5

Naïve Bayes' Classification Algorithm

Working of Naïve Bayes' Classifier

Frequency and Likelihood tables of 'Origin'

Frequency Table

		Stolen?	
		Yes	No
Origin	Domestic	2	3
	Imported	3	2

Likelihood Table

		Stolen?	
		P(Yes)	P(No)
Origin	Domestic	2/5	3/5
	Imported	3/5	2/5

So in our example, we have 3 predictors **X**.

Color	Type	Origin	Stolen
Red	SUV	Domestic	?

As per the equations discussed above, we can calculate the posterior probability P(Yes | X) and P(No | X) and can see the result

Naïve Bayes' Classification Algorithm

Solve the above problem

Naïve Bayes' Classification Algorithm

Here is a tabular representation of a dataset

Let us test it on a new set of features (let us call it today).

```
today = (Sunny, Hot, Normal, False)
```

Predict whether it is possible or not possible to play golf today

	Outlook	Temperature	Humidity	Windy	Play Golf
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

Naïve Bayes' Classification Algorithm

Example of Conditional Probability

A math teacher gave her class two tests. 25% of the class passed both tests and 42% of the class passed the first test. What percent of those who passed the first test also passed the second test?

$$P(\text{Second}|\text{First}) = \frac{P(\text{First and Second})}{P(\text{First})} = \frac{0.25}{0.42} = 0.60 = 60\%$$

Naïve Bayes' Classification Algorithm

Example of Conditional Probability

When you say the conditional probability of A given B, it denotes the probability of A occurring given that B has already occurred. Mathematically, Conditional probability of A given B can be computed as: P(A|B) = P(A AND B) / P(B)

<u>School Example</u> Let's see a slightly complicated example. Consider a school with a total population of 100 persons. These 100 persons can be seen either as 'Students' and 'Teachers' or as a population of 'Males' and 'Females'. With below tabulation of the 100 people, what is the conditional probability that a certain member of the school is a 'Teacher' given that he is a 'Man'?

Naïve Bayes' Classification Algorithm

	Female	Male	Total
Teacher	8	12	20
Student	32	48	80
Total	40	60	100

Example of Conditional Probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad (1)$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \qquad (2)$$

To calculate this, you may intuitively filter the sub-population of 60 males and focus on the 12 (male) teachers. So the required conditional probability

P(Teacher | Male) = 12 / 60 = 0.2

$$P(Teacher \mid Male) = \frac{P(Teacher \cap Male)}{P(Male)} = 12/60 = 0.2$$

Naïve Bayes' Classification Algorithm

Example of Conditional Probability

Conditional probabilities can be found simply from data in tables, as illustrated by the following.
The table opposite shows the choices of language and the gender of the 200 students choosing those languages.

	French	German	Total
Male	40	40	80
Female	90	30	120
Total	130	70	200

A student is choosing at random, find the probability of that student,

a) studying French, $\frac{130}{200}$

b) being male, $\frac{80}{200}$

c) being male and studying German, $\frac{40}{200}$

d) being female, given he/she studies French. $\frac{90}{130}$

e) studying German, given that he is male. $\frac{40}{80}$

Naïve Bayes' Classification Algorithm

Bayes' Theorem gives the conditional probability of an event A given another event B has occurred

In this case, the first coin toss will be B and the second coin toss A.

Bayes Theorem

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

where:

P(A|B) = Conditional Probability of A given B

P(B|A) = Conditional Probability of B given A

P(A) = Probability of event A

P(B) = Probability of event A

Naïve Bayes' Classification Algorithm

LET US APPLY BAYES THEOREM
TO OUR EXAMPLE

Here, the sample space is:

{HH, HT, TH, TT}

THIS USES CONDITIONAL
PROBABILITY. LET US
UNDERSTAND THIS IN DETAIL

1. P(Getting two heads) = 1/4
2. P(Atleast one tail) = 3/4
3. P(Second coin being head given first coin is tail) = 1/2
4. P(Getting two heads given first coin is a head) = 1/2

Naïve Bayes' Classification Algorithm

IN THIS SAMPLE SPACE, LET A BE THE EVENT THAT SECOND COIN IS HEAD AND B BE THE EVENT THAT FIRST COIN IS TAIL

In the sample space:

{HH, HT, TH, TT}

P(Second coin being head given first coin is tail)

= P(A|B)

= [P(B|A) * P(A)] / P(B)

= [P(First coin being tail given second coin is head) * P(Second coin being head)] / P(First coin being tail)

= [(1/2) * (1/2)] / (1/2)

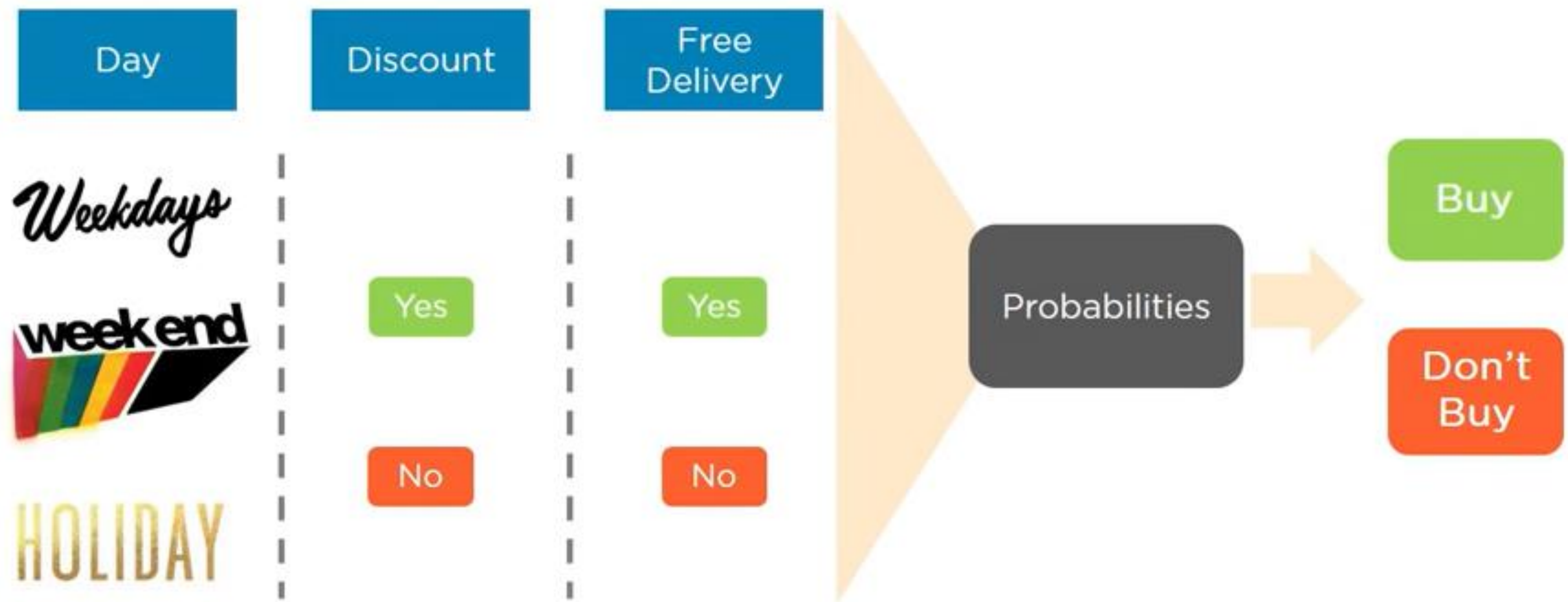
= 1/2 = 0.5

We're going to focus on A, and we write that out as a probability of A given B

Naïve Bayes' Classification Algorithm

Shopping Demo – Problem Statement

To predict whether a person will purchase a product on a specific combination of **Day**, **Discount** and **Free Delivery** using **Naive Bayes Classifier**



Naïve Bayes' Classification Algorithm

Shopping Demo – Dataset

We have a small sample dataset of 30 rows for our demo

	A	B	C	D
1	Day	Discount	Free Delivery	Purchase
2	Weekday	Yes	Yes	Yes
3	Weekday	Yes	Yes	Yes
4	Weekday	No	No	No
5	Holiday	Yes	Yes	Yes
6	Weekend	Yes	Yes	Yes
7	Holiday	No	No	No
8	Weekend	Yes	No	Yes
9	Weekday	Yes	Yes	Yes
10	Weekend	Yes	Yes	Yes
11	Holiday	Yes	Yes	Yes
12	Holiday	No	Yes	Yes
13	Holiday	No	No	No
14	Weekend	Yes	Yes	Yes
15	Holiday	Yes	Yes	Yes

Naive_Bayes_Dataset

Naïve Bayes' Classification Algorithm

Shopping Demo – Frequency Table

Based on this dataset containing three input types of *Day*, *Discount* and *Free Delivery*, we will populate frequency tables for each attribute

Frequency Table		Buy	
		Yes	No
Discount	Yes	19	1
	No	5	5

Frequency Table		Buy	
		Yes	No
Free Delivery	Yes	21	2
	No	3	4

Frequency Table		Buy	
		Yes	No
Day	Weekday	9	2
	Weekend	7	1
	Holiday	8	3

Naïve Bayes' Classification Algorithm

Shopping Demo – Frequency Table

Based on this dataset containing three input types of *Day*, *Discount* and *Free Delivery*, we will populate frequency tables for each attribute

Frequency Table		Buy	
		Yes	No
Discount	Yes	19	1
	No	5	5

Frequency Table		Buy	
		Yes	No
Free Delivery	Yes	21	2
	No	3	4

Frequency Table		Buy	
		Yes	No
Day	Weekday	9	2
	Weekend	7	1
	Holiday	8	3

A

B

FOR OUR BAYES THEOREM, LET THE EVENT BUY BE A AND THE INDEPENDENT VARIABLES, DISCOUNT, FREE DELIVERY AND DAY BE B

Naïve Bayes’ Classification Algorithm

Shopping Demo – Likelihood Table

Now let us calculate the Likelihood table for one of the variable, *Day* which includes ***Weekday***, ***Weekend*** and ***Holiday***

Frequency Table		Buy: Yes	Buy: No	
Day	Weekday	9	2	11
	Weekend	7	1	8
	Holiday	8	3	11
		24	6	30

Likelihood Table		Buy: Yes	Buy: No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

P(B) = **P(Weekday)**
= 11/30 = 0.37

P(A) = **P(No Buy)**
= 6/30 = 0.2

P(B|A)
= **P(Weekday | No Buy)**
= 2/6 = 0.33

Naïve Bayes' Classification Algorithm

Shopping Demo – Likelihood Table

Based on this likelihood table, we will calculate conditional probabilities as below

Frequency Table		Buy: Yes	Buy: No	
Day	Weekday	9	2	11
	Weekend	7	1	8
	Holiday	8	3	11
		24	6	30

Likelihood Table		Buy: Yes	Buy: No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

P(B) = **P(Weekday)** = 11/30 = 0.367

P(A) = **P(No Buy)** = 6/30 = 0.2

P(B|A) = **P(Weekday | No Buy)** = 2/6 = 0.33

P (A|B) = **P (No Buy | Weekday)**
= P(Weekday| No Buy) * P(No Buy) / P(Weekday)
= (0.33 * 0.2) / 0.367 = 0.179

As the **Probability(Buy | Weekday)** is more than **Probability(No Buy | Weekday)**, we can conclude that a customer will most likely buy the product on a Weekday

Naïve Bayes' Classification Algorithm

Shopping Demo – Naive Bayes Classifier

Similarly, we can find the likelihood of occurrence of an event involving all three variables

WE HAVE THE FREQUENCY TABLES OF ALL THE THREE INDEPENDENT VARIABLES. WE WILL NOW CONSTRUCT LIKELIHOOD TABLES FOR ALL THE THREE

Frequency Table		Buy	
		Yes	No
Discount	Yes	9	2
	No	5	14

Frequency Table		Buy	
		Yes	No
Free Delivery	Yes	6	3
	No	5	16

Frequency Table		Buy	
		Yes	No
Day	Weekday	9	2
	Weekend	7	1
	Holiday	8	3

Naïve Bayes' Classification Algorithm

Shopping Demo – Naive Bayes Classifier

Similarly, we can find the likelihood of occurrence of an event involving all three variables

WE HAVE THE FREQUENCY TABLES OF ALL THE THREE INDEPENDENT VARIABLES. WE WILL NOW CONSTRUCT LIKELIHOOD TABLES FOR ALL THE THREE

Frequency Table		Buy	
		Yes	No
Day	Weekday	3	7
	Weekend	8	2
	Holiday	9	1

Likelihood Table		Buy		
		Yes	No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

Naïve Bayes' Classification Algorithm

Shopping Demo – Naive Bayes Classifier

Similarly, we can find the likelihood of occurrence of an event involving all three variables

WE HAVE THE FREQUENCY TABLES OF ALL THE THREE INDEPENDENT VARIABLES. WE WILL NOW CONSTRUCT LIKELIHOOD TABLES FOR ALL THE THREE

Frequency Table		Buy	
		Yes	No
Discount	Yes	19	1
	No	5	5

Frequency Table		Buy		
		Yes	No	
Discount	Yes	19/24	1/6	20/30
	No	5/24	5/6	10/30
		24/30	6/30	

Naïve Bayes' Classification Algorithm

Shopping Demo – Naive Bayes Classifier

LET US USE THESE 3 LIKELIHOOD TABLES TO CALCULATE WHETHER A CUSTOMER WILL PURCHASE A PRODUCT ON A SPECIFIC COMBINATION OF DAY, DISCOUNT AND FREE DELIVERY OR NOT

HERE, LET US TAKE A COMBINATION OF THESE FACTORS:

- DAY = HOLIDAY
- DISCOUNT = YES
- FREE DELIVERY = YES

Naïve Bayes' Classification Algorithm

Shopping Demo – No Purchase

Likelihood Tables

Likelihood Table		Buy: Yes	Buy: No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

Frequency Table		Buy: Yes	Buy: No	
Discount	Yes	19/24	1/6	20/30
	No	5/24	5/6	10/30
		24/30	6/30	

Frequency Table		Buy: Yes	Buy: No	
Free Delivery	Yes	21/24	2/6	23/30
	No	3/24	4/6	7/30
		24/30	6/30	

Calculating Conditional Probability of purchase on the following combination of day, discount and free delivery:

Where B equals:

- Day = Holiday
- Discount = Yes
- Free Delivery = Yes

Let A = No Buy

P (A|B) = P (No Buy | Discount = Yes, Free Delivery = Yes, Day = Holiday)

$$= \frac{P(\text{Discount} = \text{Yes} \mid \text{No}) * P(\text{Free Delivery} = \text{Yes} \mid \text{No}) * P(\text{Day} = \text{Holiday} \mid \text{No}) * P(\text{No Buy})}{P(\text{Discount=Yes}) * P(\text{Free Delivery=Yes}) * P(\text{Day=Holiday})}$$

$$= \frac{(1/6) * (2/6) * (3/6) * (6/30)}{(20/30) * (23/30) * (11/30)}$$

$$= 0.178$$

Naïve Bayes' Classification Algorithm

Shopping Demo – Purchase

Likelihood Tables

Likelihood Table		Buy: Yes	Buy: No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

Frequency Table		Buy: Yes	Buy: No	
Discount	Yes	19/24	1/6	20/30
	No	5/24	5/6	10/30
		24/30	6/30	

Frequency Table		Buy: Yes	Buy: No	
Free Delivery	Yes	21/24	2/6	23/30
	No	3/24	4/6	7/30
		24/30	6/30	

Calculating Conditional Probability of purchase on the following combination of day, discount and free delivery:

Where B equals:

- Day = Holiday
- Discount = Yes
- Free Delivery = Yes

Let A = Buy

P (A|B) = P (Yes Buy | Discount = Yes, Free Delivery = Yes, Day = Holiday)

$$= \frac{P(\text{Discount} = \text{Yes} \mid \text{Yes}) * P(\text{Free Delivery} = \text{Yes} \mid \text{Yes}) * P(\text{Day} = \text{Holiday} \mid \text{Yes}) * P(\text{Yes Buy})}{P(\text{Discount}=\text{Yes}) * P(\text{Free Delivery}=\text{Yes}) * P(\text{Day}=\text{Holiday})}$$

$$= \frac{(19/24) * (21/24) * (8/24) * (24/30)}{(20/30) * (23/30) * (11/30)}$$

$$= 0.986$$

Naïve Bayes' Classification Algorithm

Shopping Demo – Naive Bayes Classifier

PROBABILITY OF PURCHASE = 0.986
PROBABILITY OF NO PURCHASE = 0.178

LET US NOW NORMALIZE THESE PROBABILITIES TO GET THE LIKELIHOOD OF THE EVENTS

FINALLY, WE HAVE CONDITIONAL PROBABILITIES OF PURCHASE ON THIS DAY!

Naïve Bayes’ Classification Algorithm

Shopping Demo – Result

SUM OF PROBABILITIES
= 0.986 + 0.178 = 1.164

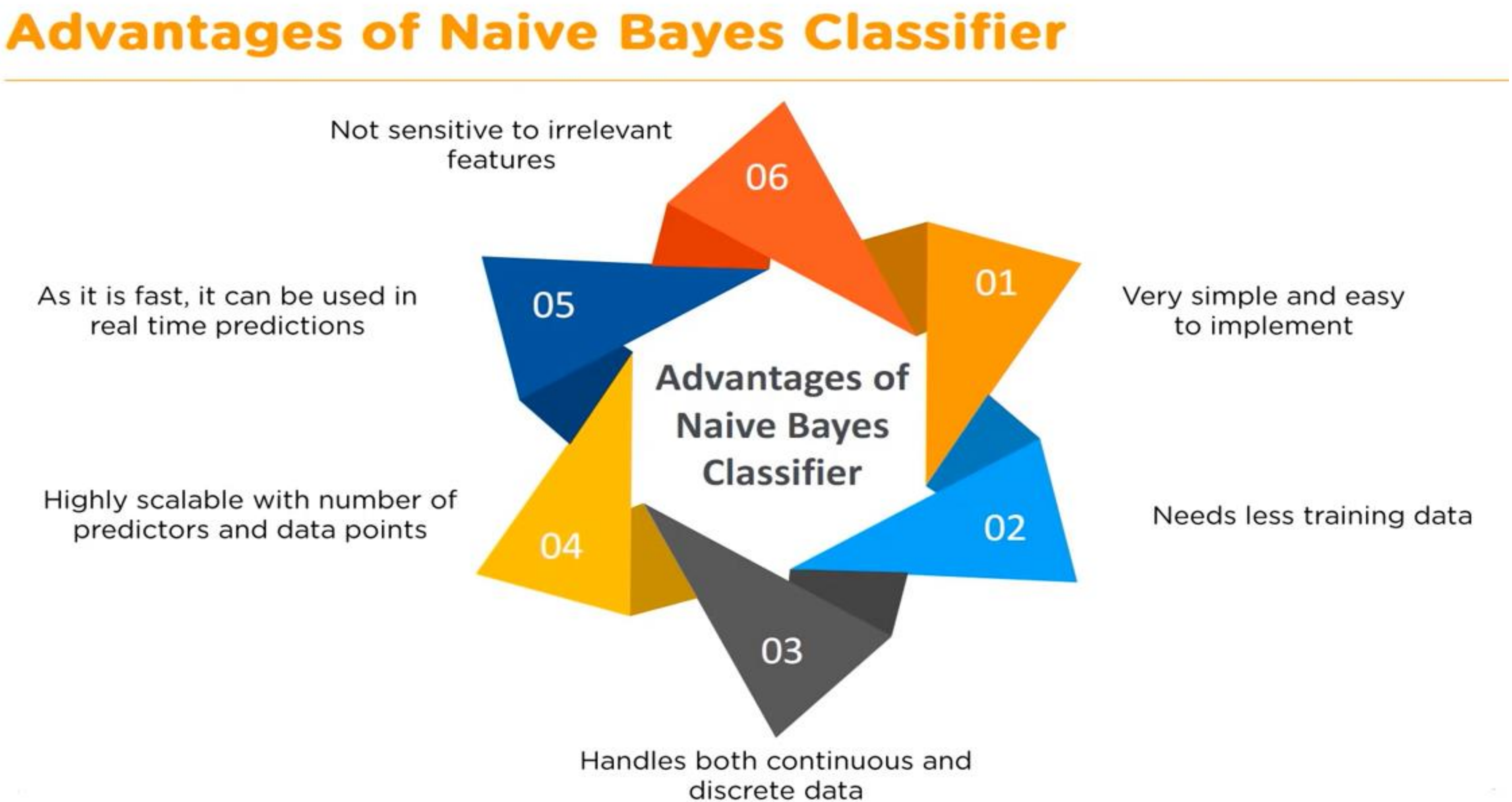
LIKELIHOOD OF **PURCHASE**
= 0.986 / 1.164 = **84.71 %**

LIKELIHOOD OF **NO PURCHASE**
= 0.178 / 1.164 = **15.29 %**

PROBABILITY OF PURCHASE = 0.986
PROBABILITY OF NO PURCHASE = 0.178

AS **84.71%** IS GREATER THAN **15.29%**, WE CAN CONCLUDE THAT AN AVERAGE CUSTOMER WILL **BUY** ON A **HOLIDAY WITH DISCOUNT AND FREE DELIVERY**

Naïve Bayes' Classification Algorithm



Naïve Bayes' Classification Algorithm

Naive Bayesian

The Naive Bayesian classifier is based on Bayes' theorem with the independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

Algorithm

Bayes theorem provides a way of calculating the posterior probability, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

- $P(c|x)$ is the posterior probability of *class* (*target*) given *predictor* (*attribute*).
- $P(c)$ is the prior probability of *class*.
- $P(x|c)$ is the likelihood which is the probability of *predictor* given *class*.
- $P(x)$ is the prior probability of *predictor*.

Naïve Bayes' Classification Algorithm

Example 1:

We use the same simple Weather dataset here.

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

The posterior probability can be calculated by first, constructing a frequency table for each attribute against the target. Then, transforming the frequency tables to likelihood tables and finally use the Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

Naïve Bayes' Classification Algorithm

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

$$P(x \mid c) = P(Sunny \mid Yes) = 3/9 = 0.33$$

Likelihood Table		Play Golf		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	2/9	3/5	5/14
		9/14	5/14	

$$P(x) = P(Sunny) = 5/14 = 0.36$$

$$P(c) = P(Yes) = 9/14 = 0.64$$

Posterior Probability: $P(c \mid x) = P(Yes \mid Sunny) = 0.33 \times 0.64 \div 0.36 = 0.60$

Naïve Bayes' Classification Algorithm

$$P(x \mid c) = P(Sunny \mid No) = 2/5 = 0.4$$

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
		9	5	14

$$P(x) = P(Sunny) = 5/14 = 0.36$$

$$P(c) = P(No) = 5/14 = 0.36$$

Posterior Probability: $P(c \mid x) = P(No \mid Sunny) = 0.40 \times 0.36 \div 0.36 = 0.40$

Naïve Bayes' Classification Algorithm

The likelihood tables for all four predictors.

Frequency Table

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3

Likelihood Table

		Play Golf	
		Yes	No
Outlook	Sunny	3/9	2/5
	Overcast	4/9	0/5
	Rainy	2/9	3/5

		Play Golf	
		Yes	No
Humidity	High	3/9	4/5
	Normal	6/9	1/5

		Play Golf	
		Yes	No
Temp.	Hot	2/9	2/5
	Mild	4/9	2/5
	Cool	3/9	1/5

		Play Golf	
		Yes	No
Windy	False	6/9	2/5
	True	3/9	3/5

Naïve Bayes' Classification Algorithm

Example 2:

In this example we have 4 inputs (predictors). The final posterior probabilities can be standardized between 0 and 1.

Outlook	Temp	Humidity	Windy	Play
Rainy	Cool	High	True	?

$$P(Yes \mid X) = P(Rainy \mid Yes) \times P(Cool \mid Yes) \times P(High \mid Yes) \times P(True \mid Yes) \times P(Yes)$$

$$P(Yes \mid X) = 2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.00529 \rightarrow 0.2 = \frac{0.00529}{0.02057 + 0.00529}$$

$$P(No \mid X) = P(Rainy \mid No) \times P(Cool \mid No) \times P(High \mid No) \times P(True \mid No) \times P(No)$$

$$P(No \mid X) = 3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.02057 \rightarrow 0.8 = \frac{0.02057}{0.02057 + 0.00529}$$

Thank you