



Machine Learning

Phd. César Astudillo | Facultad de Ingeniería

Naïve Bayes

Content

- What is Naive Bayes?
- Understanding Naive Bayes and Machine Learning
- Where is Naive Bayes Used?
- Understanding Naive Bayes Classifier
- Advantages of Naive Bayes
- Classifier Use Case - Text Classification



Introducing Naive Bayes Classifier

HAVE YOU EVER WONDERED HOW
YOUR MAIL PROVIDER
IMPLEMENTS SPAM FILTERING?



Introducing Naive Bayes Classifier

HAVE YOU EVER WONDERED HOW
YOUR MAIL PROVIDER
IMPLEMENTS SPAM FILTERING?



OR HOW ONLINE NEWS
CHANNELS PERFORM NEWS
TEXT CLASSIFICATION?



Introducing Naive Bayes Classifier

HAVE YOU EVER WONDERED HOW
YOUR MAIL PROVIDER
IMPLEMENTS SPAM FILTERING?

OR HOW ONLINE NEWS
CHANNELS PERFORM NEWS
TEXT CLASSIFICATION?

OR HOW COMPANIES PERFORM
SENTIMENT ANALYSIS OF THEIR
AUDIENCE ON SOCIAL MEDIA?



Introducing Naive Bayes Classifier

ALL OF THIS AND MORE IS
DONE THROUGH A MACHINE
LEARNING ALGORITHM CALLED
NAIVE BAYES CLASSIFIER



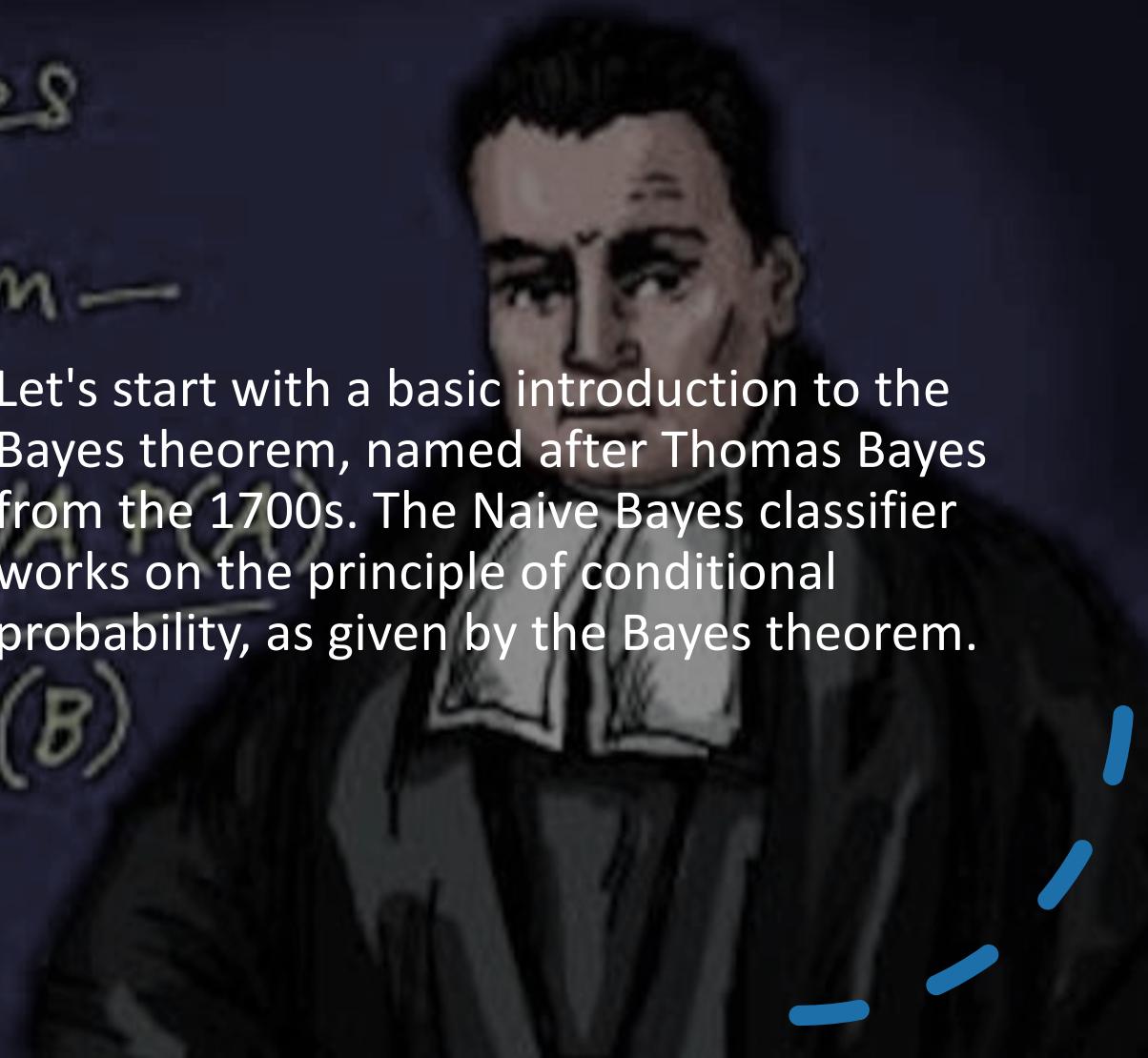
What is Naïve Bayes?

Thomas Bayes

Bayes' theorem —

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

Let's start with a basic introduction to the Bayes theorem, named after Thomas Bayes from the 1700s. The Naive Bayes classifier works on the principle of conditional probability, as given by the Bayes theorem.



Introducing Bayes Theorem



NAIVE BAYES CLASSIFIER
WORKS ON THE PRINCIPLES OF
CONDITIONAL PROBABILITY AS
GIVEN BY THE BAYES' THEOREM

Introducing Bayes Theorem



LET US CONSIDER THE
FOLLOWING EXAMPLE OF
TOSSING TWO COINS



Here, the sample space is:

$$\{HH, HT, TH, TT\}$$

1. $P(\text{Getting two heads}) = 1/4$
2. $P(\text{At least one tail}) = 3/4$
3. $P(\text{Second coin being head given first coin is tail}) = 1/2$
4. $P(\text{Getting two heads given first coin is a head}) = 1/2$

Introducing Bayes Theorem

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

Bayes Theorem

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

where:

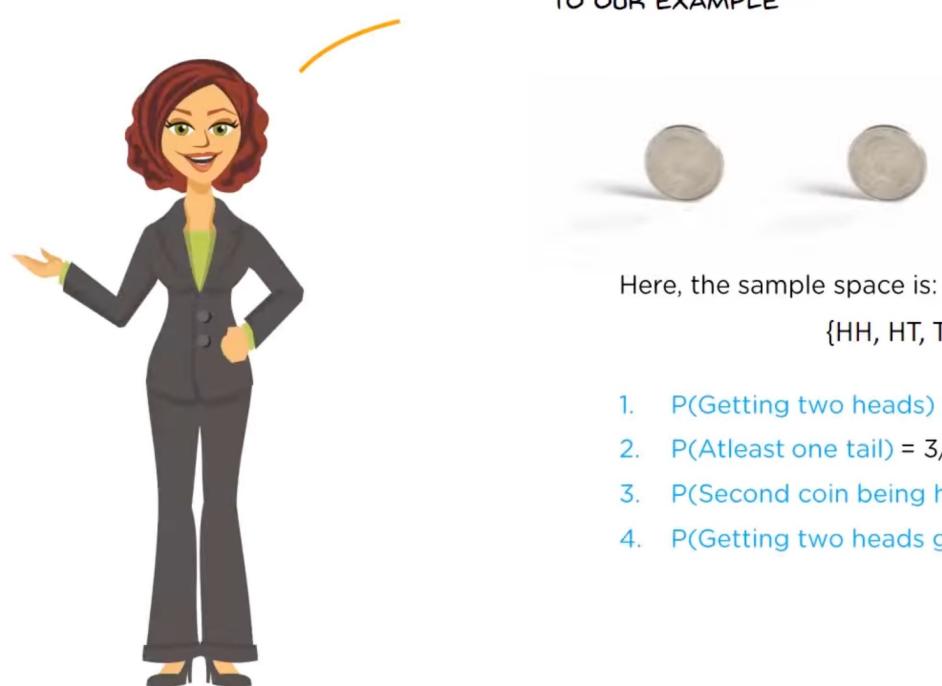
$P(A|B)$ = Conditional Probability of A given B

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

$P(A)$ = Probability of event A

$P(B)$ = Probability of event B

Introducing Bayes Theorem



LET US APPLY BAYES THEOREM
TO OUR EXAMPLE



Here, the sample space is:

$$\{HH, HT, TH, TT\}$$

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

Introducing Bayes Theorem



LET US APPLY BAYES THEOREM
TO OUR EXAMPLE



Here, the sample space is:

$$\{HH, HT, TH, TT\}$$

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

THESE TWO USE SIMPLE
PROBABILITIES CALCULATED DIRECTLY
FROM THE SAMPLE SPACE

Introducing Bayes Theorem



LET US APPLY BAYES THEOREM
TO OUR EXAMPLE



Here, the sample space is:

$$\{HH, HT, TH, TT\}$$

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

THIS USES CONDITIONAL
PROBABILITY. LET US
UNDERSTAND THIS IN DETAIL

Introducing Naive Bayes Classifier



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:

$$\{HH, HT, TH, TT\}$$

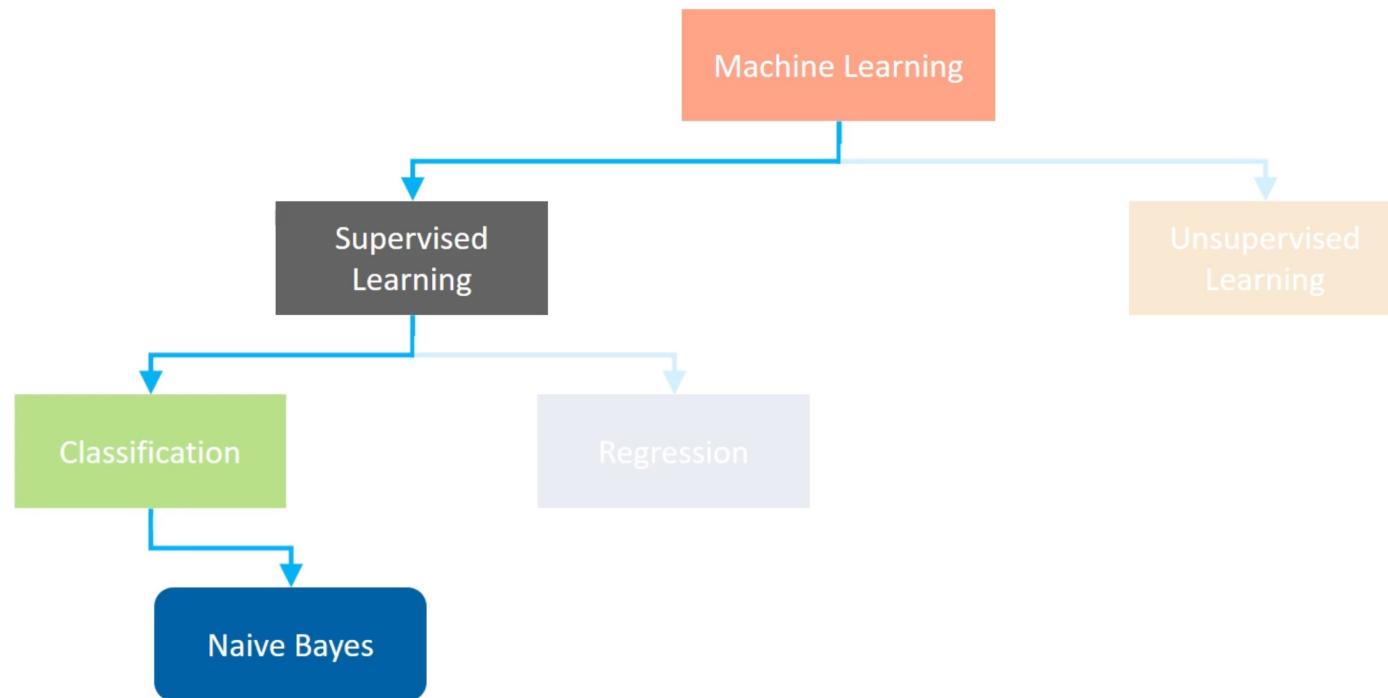
$$\begin{aligned} & P(\text{Second coin being head given first coin is tail}) \\ &= P(A|B) \\ &= [P(B|A) * P(A)] / P(B) \\ &= [P(\text{First coin being tail given second coin is head}) * P(\text{Second coin being head})] / P(\text{First coin being tail}) \\ &= [(1/2) * (1/2)] / (1/2) \\ &= 1/2 = 0.5 \end{aligned}$$

Introducing Bayes Theorem



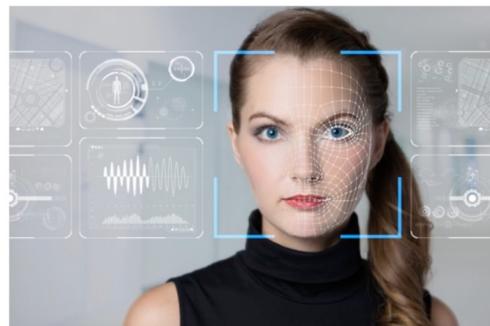
BAYES' THEOREM BASICALLY CALCULATES
THE CONDITIONAL PROBABILITY OF THE
OCCURRENCE OF AN EVENT BASED ON
PRIOR KNOWLEDGE OF CONDITIONS THAT
MIGHT BE RELATED TO THE EVENT

Understanding Naive Bayes and Machine Learning



Where is Naive Bayes used?

Face
Recognition



Weather
Prediction



Where is Naive Bayes used?

Medical
Diagnosis



News
Classification

A screenshot of the Google News interface. The top bar says "Google News". Below it, under the heading "World", there are several news stories: "Easter 2018: Here's how the Easter Sunday date is determined" (The Indian Express, 1 hr ago), "Check out Easter festivities at the Wigwam" (ABC13 Houston, 40m ago), "Ever wonder why the day Jesus Christ died is called 'Good Friday'?" (India Today, 1 hr ago), "Good Friday 2018: Christians take out processions, observe fast on day marking crucifixion of Jesus Christ [Photos]" (Forbes, 1 hr ago), and "Jesus is the greatest mystery told" (In-depth - The Australian, 57m ago). There is a link to "View full coverage →". On the left, a sidebar shows "SECTIONS" with links to Top Stories, India, World, Business, Technology, Entertainment, Sport, Science, and Health. On the right, a "Related" sidebar lists Gaza Strip, Cambridge Analytica, Malala Yousafzai, Israel, Pakistan, Hamas, Facebook, Syria, Palestinians, and Good Friday.

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

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

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

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

where:

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

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

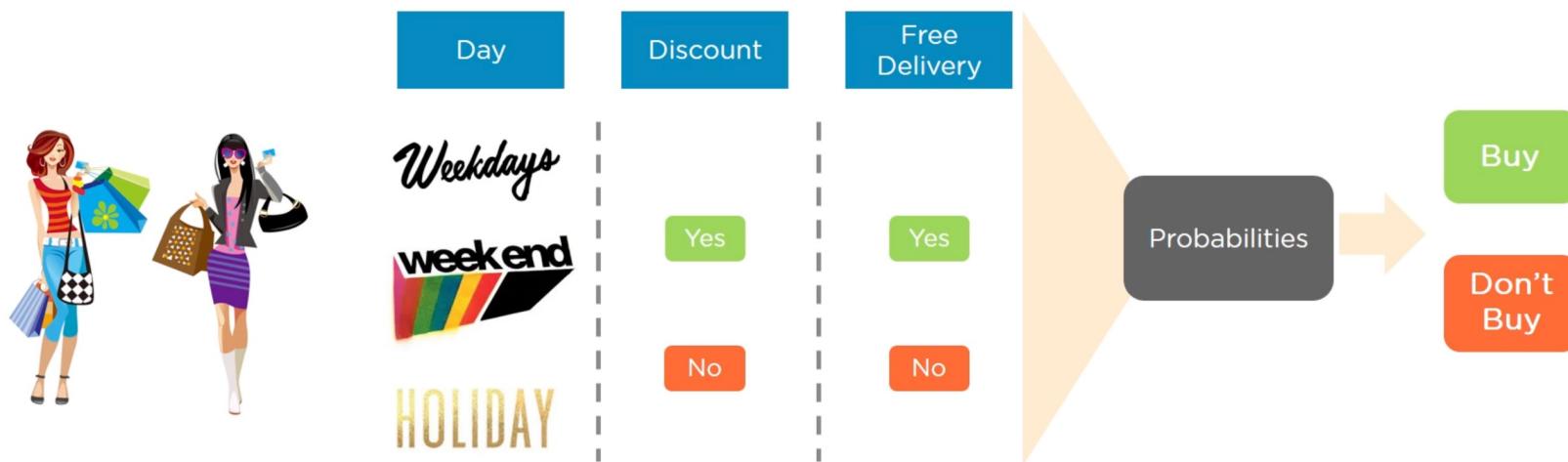
$P(A)$ = Probability of event A

$P(B)$ = Probability of event B

Case Study: Shopping Demo

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



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 

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

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

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



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

Likelihood Table		Buy		
Day	Weekday	Yes	No	
	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	

$$P(B) = P(\text{Weekday}) \\ = 11/30 = 0.37$$

$$P(A) = P(\text{No Buy}) \\ = 6/30 = 0.2$$

$$P(B|A) \\ = P(\text{Weekday} | \text{No Buy}) \\ = 2/6 = 0.33$$

Shopping Demo – Likelihood Table

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

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

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

$$P(B) = P(\text{Weekday}) = 11/30 = 0.367$$

$$P(A) = P(\text{No Buy}) = 6/30 = 0.2$$

$$P(B|A) = P(\text{Weekday} \mid \text{No Buy}) = 2/6 = 0.33$$

$$P(A|B) = P(\text{No Buy} \mid \text{Weekday})$$

$$= P(\text{Weekday} \mid \text{No Buy}) * P(\text{No Buy}) / P(\text{Weekday})$$

$$= (0.33 * 0.2) / 0.367 = 0.179$$

Shopping Demo – Likelihood Table

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

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

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	

$$P(B) = P(\text{Weekday}) = 11/30 = 0.367$$

$$P(A) = P(\text{Buy}) = 24/30 = 0.8$$

$$P(B|A) = P(\text{Weekday} | \text{Buy}) = 2/6 = 0.375$$

If A equals Buy, then

$$P(A|B) = P(\text{Buy} | \text{Weekday})$$

$$= P(\text{Weekday} | \text{Buy}) * P(\text{Buy}) / P(\text{Weekday})$$

$$= (0.375 * 0.8) / 0.367 = 0.817$$

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

Shopping Demo – Naive Bayes Classifier

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

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

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



Shopping Demo – Naive Bayes Classifier

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

Frequency Table		Buy	
Discount	Yes	19	1
	No	5	5

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

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



Shopping Demo – Naive Bayes Classifier

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

Frequency Table		Buy	
Free Delivery	Yes	21	2
	No	3	4

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

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



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

Shopping Demo – No Purchase

Likelihood Tables

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	

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

Frequency Table		Buy		
		Yes	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

$$P(A|B) = P(\text{No Buy} | \text{Discount} = \text{Yes}, \text{Free Delivery} = \text{Yes}, \text{Day} = \text{Holiday})$$

$$\begin{aligned}
 &= \frac{P(\text{Discount} = \text{Yes} | \text{No}) * P(\text{Free Delivery} = \text{Yes} | \text{No}) * P(\text{Day} = \text{Holiday} | \text{No}) * P(\text{No Buy})}{P(\text{Discount}=\text{Yes}) * P(\text{Free Delivery}=\text{Yes}) * P(\text{Day}=\text{Holiday})} \\
 &= \frac{(1/6) * (2/6) * (3/6) * (6/30)}{(20/30) * (23/30) * (11/30)} \\
 &= 0.178
 \end{aligned}$$

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

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

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 B

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

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

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 B

Shopping Demo – Naive Bayes Classifier

PROBABILITY OF PURCHASE = **0.986**
PROBABILITY OF NO PURCHASE = **0.178**

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



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

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

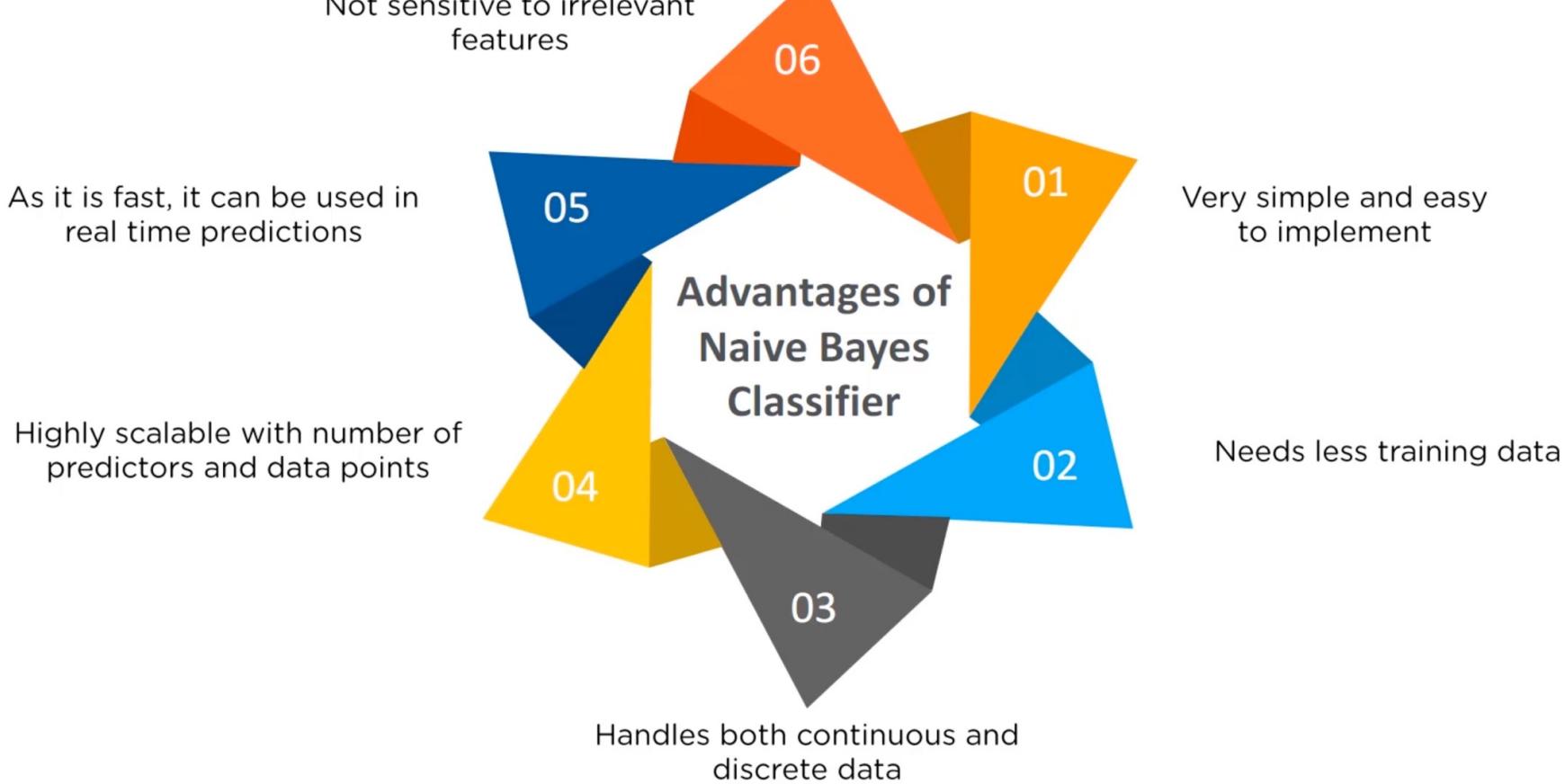


AS 84.71% IS GREATER THAN 15.29%,
WE CAN CONCLUDE THAT AN AVERAGE
CUSTOMER WILL BUY ON A HOLIDAY
WITH DISCOUNT AND FREE
DELIVERY



Advantages of the Naïve Bayes Classifier

Advantages of Naive Bayes Classifier





Machine Learning

Phd. César Astudillo | Facultad de Ingeniería