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?

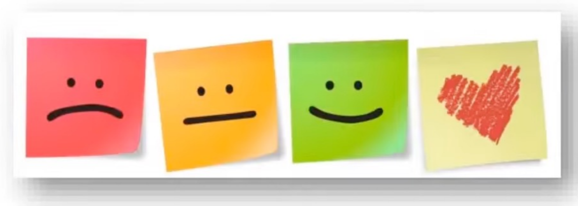
OR HOW ONLINE NEWS CHANNELS PERFORM NEWS TEXT CLASSIFICATION?

Introducing Naive Bayes Classifier

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

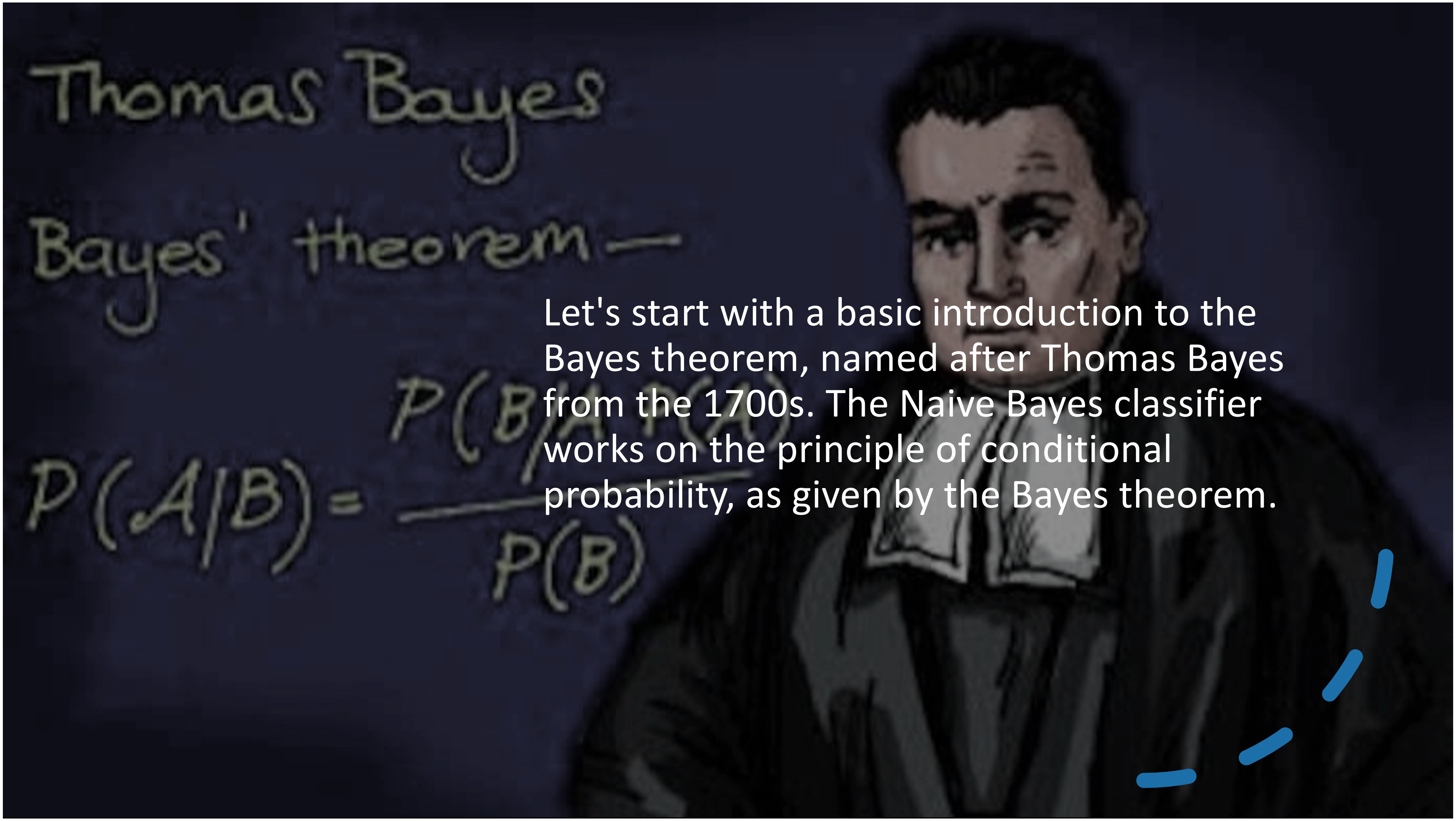
OR HOW ONLINE NEWS
CHANNELS PERFORM NEWS
TEXT CLASSIFICATION?

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



Introducing Naive Bayes Classifier

What is Naïve Bayes?



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

Introducing Bayes Theorem

Here, the sample space is:

{HH, HT, TH, TT}

1. P(Getting two heads) = 1/4
2. P(At least 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

Introducing Bayes Theorem

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

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

Introducing Bayes Theorem

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

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

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 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)

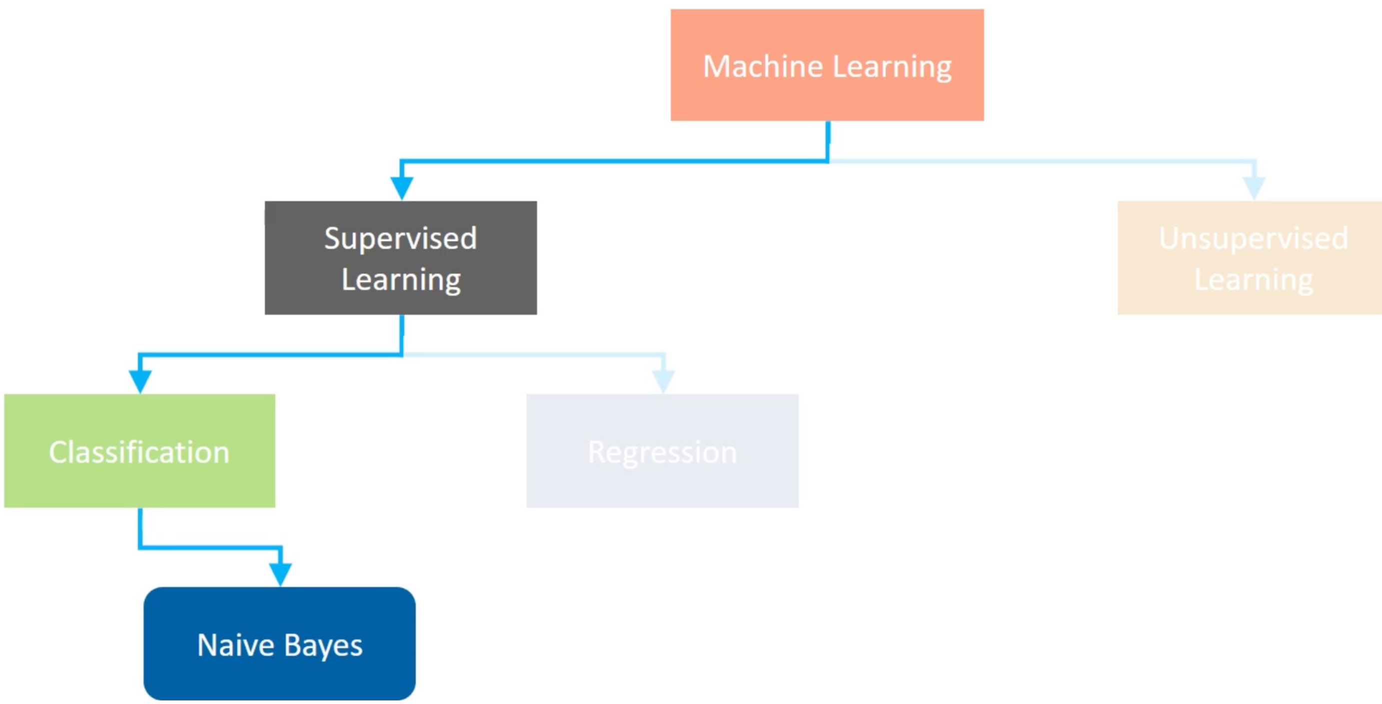
= [(1/2) * (1/2)] / (1/2)

= 1/2 = 0.5

Introducing Bayes Theorem

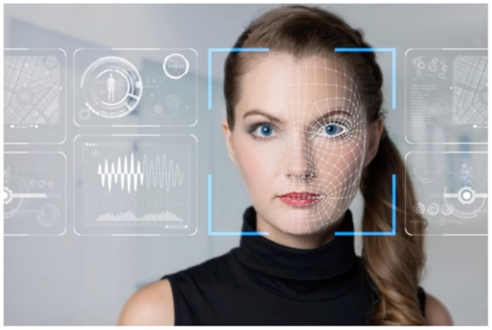
BAYES' THEOREM BASICALLY CALCULATES THE CONDITIONAL PROBABILITY OF THE OCCURRENCE OF AN EVENT BASED ON PRIOR KNOWLEDGE OF CONDITIONS THAT MIGHT BE RELATED TO THE EVENT

Understanding Naive Bayes and Machine Learning



Where is Naive Bayes used?

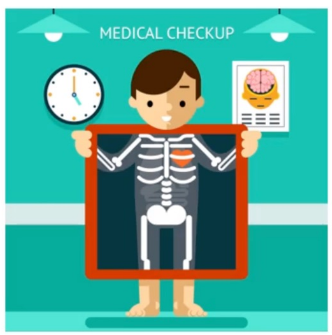
Face Recognition



Weather Prediction

Where is Naive Bayes used?

Medical Diagnosis



News Classification

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

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

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

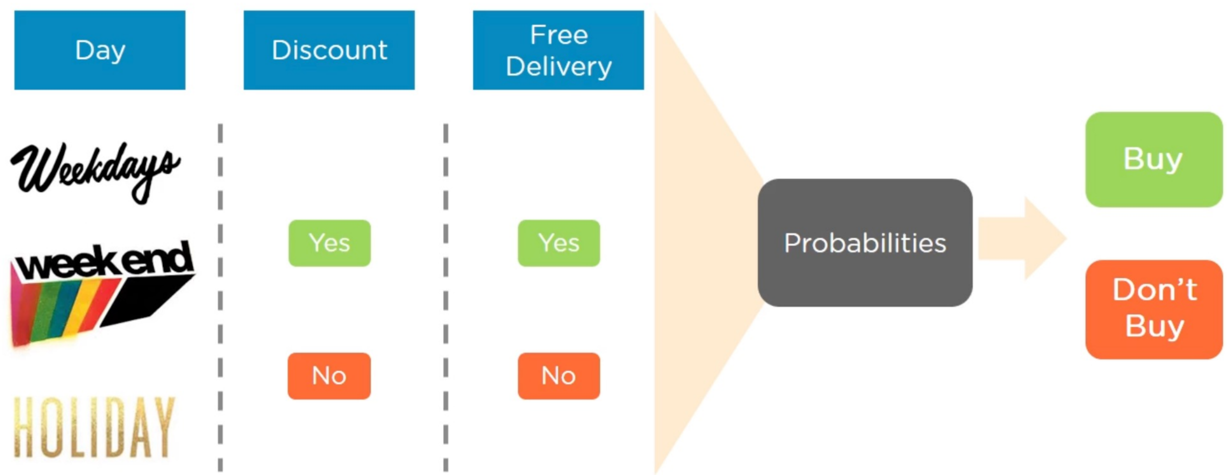
$P(A)$ = Probability of event A

$P(B)$ = Probability of event A

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	Buy: No
Discount	Yes	19	1
	No	5	5

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

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

A

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: 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

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

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(Weekday)** = 11/30 = 0.367

P(A) = **P(Buy)** = 24/30 = 0.8

P(B|A) = **P(Weekday | Buy)** = 2/6 = 0.375

If A equals Buy, then

P (A|B) = **P (Buy | Weekday)**
= P(Weekday| Buy) * P(Buy) / P(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	Buy: No
Discount	Yes	9	2
	No	5	14

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

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

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	

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
Free Delivery	Yes	21	2
	No	3	4

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	

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

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

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

$$= 0.178$$

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

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

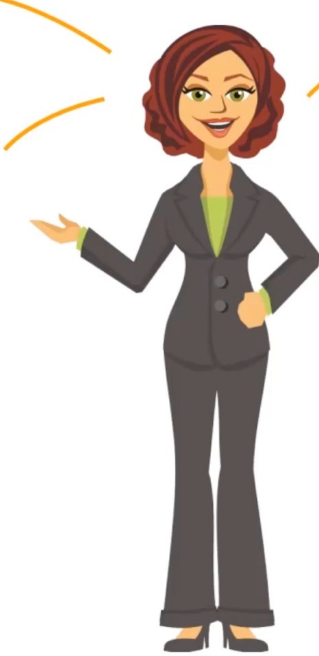
$P(A)$ = Probability of event A

$P(B)$ = Probability of event A

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 %**

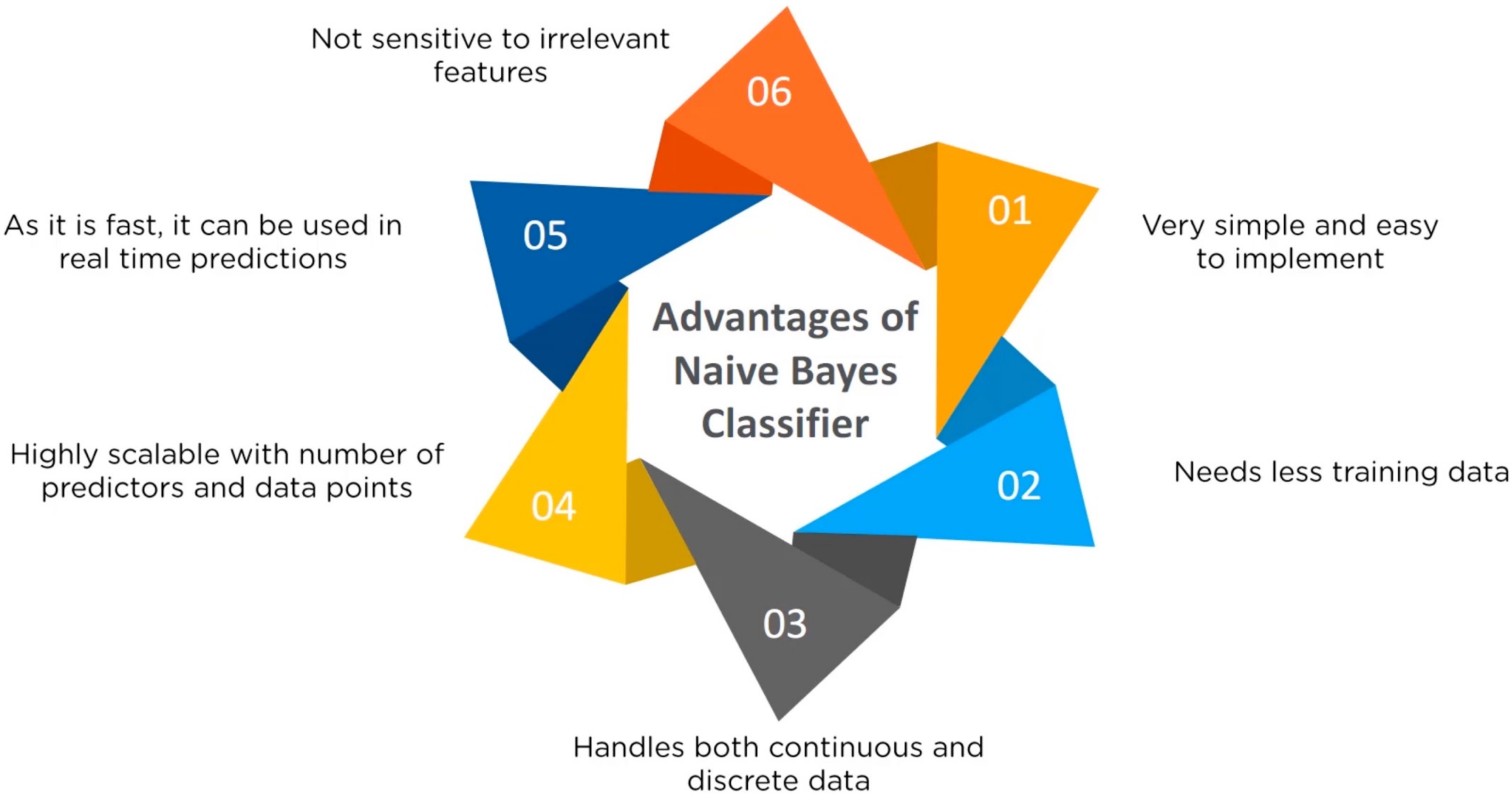
LIKELIHOOD OF **NO PURCHASE**
= 0.178 / 1.164 = **15.29 %**

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