Naïve Bayes Algorithm

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

BEFORE WE MOVE AHEAD, LET US GO THROUGH SOME OF THE SIMPLE CONCEPTS IN PROBABILITY THAT WE WILL BE USING

LET US CONSIDER THE FOLLOWING EXAMPLE OF TOSSING TWO COINS

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

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

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

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

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)

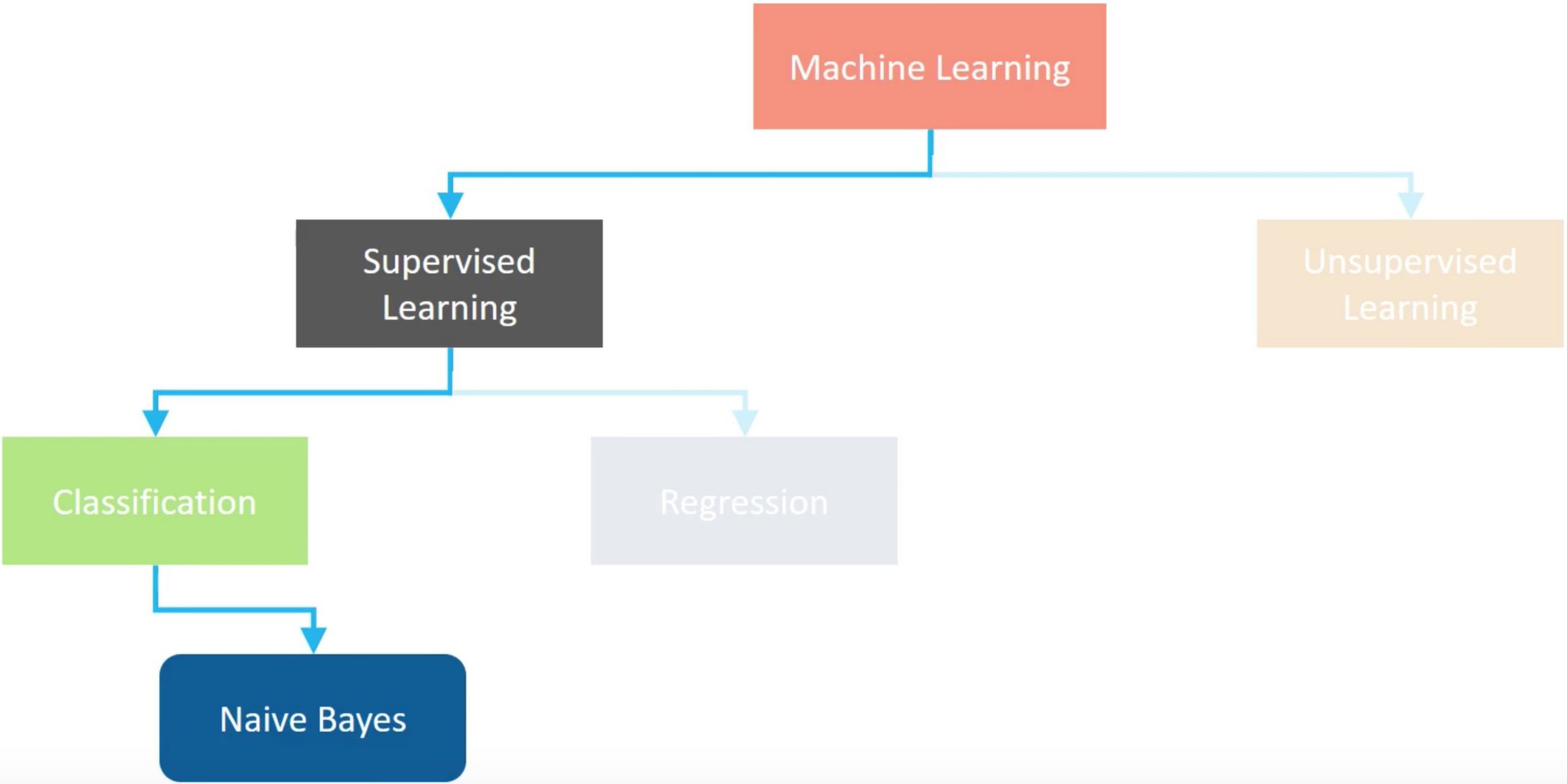
= [P(B|A) * P(A)] / P(B)

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

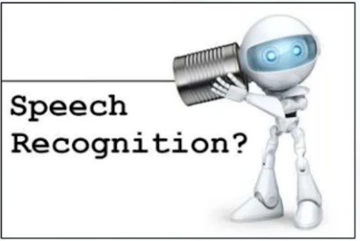
= [(1/2) * (1/2)] / (1/2)

= 1/2 = 0.5

Understanding Naïve Bayes Classifier


Machine Learning
Supervised Learning
Unsupervised Learning
Classification
Regression
Naive Bayes

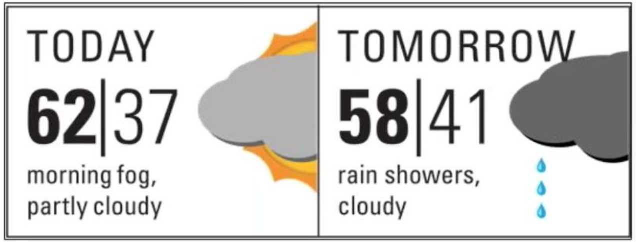
Where is it used?



Speech Recognition

Face Recognition

Anti Virus



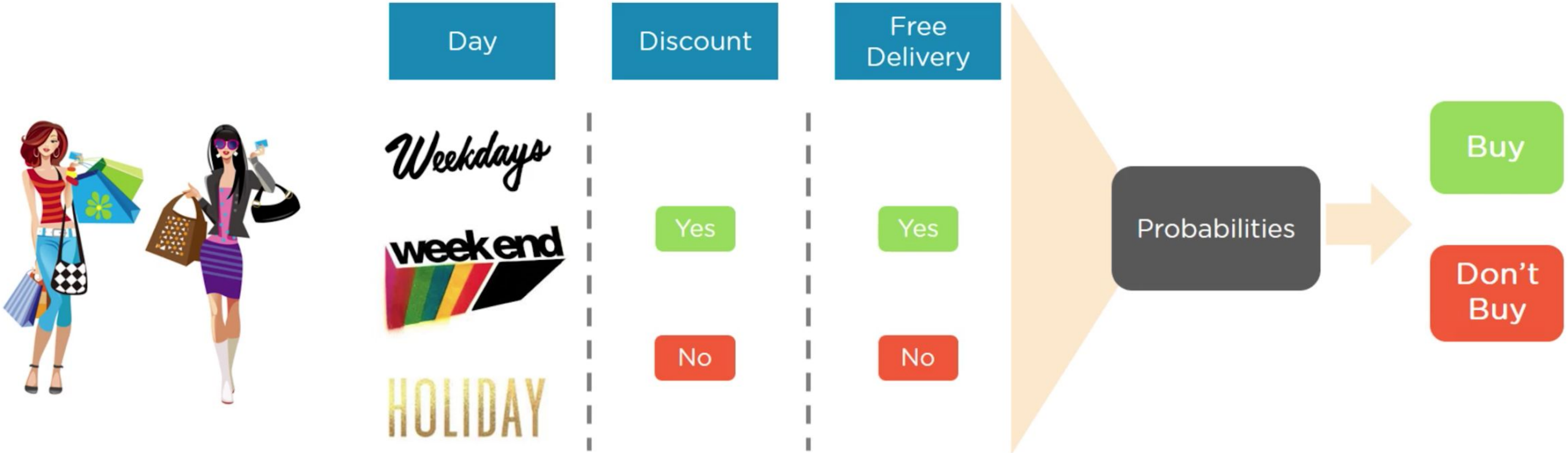
Weather Prediction

Let us understand how Bayes' Theorem can be used in Naive Bayes classifier:

$$\underbrace{P(c|x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x|c)}^{\text{Likelihood}}\;\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

Let us learn with an Example - 1

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



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

Converting it to frequency table on each category

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

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

Creating a 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	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 = 0.37

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

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

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(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

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(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

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

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

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)

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

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

$$= 0.178$$

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)

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

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

$$= 0.986$$

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

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**

Example 2

From the dataset we have obtained, we will populate frequency tables for each of the attribute

Frequency Table		Play	
		Yes	No
Sunny	Yes	3	4
	No	6	1

Frequency Table		Play	
		Yes	No
Windy	Yes	6	2
	No	3	3

Frequency Table		Play	
		Yes	No
Season	Summer	3	2
	Monsoon	4	0
	Winter	2	3

For each of the frequency tables, we will find the likelihoods for each of the cases

Here, *c = Play* and *x = Variables like Season, Sunny & Windy*.

Frequency Table		Play: Yes	Play: No
Season	Summer	3	2
	Monsoon	4	0
	Winter	2	3

Likelihood Table		Play: Yes	Play: No	
Season	Summer	3/9	2/5	5/14
	Monsoon	4/9	0/5	4/14
	Winter	2/9	3/5	5/14
		9/14	5/14	

P(x | c) = P(Summer | Yes)
= 3/9 = 0.33

P(x) = P(Summer)
= 5/14 = 0.36

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

Likelihood of 'Yes' given Summer is:

P(c | x) = P(Yes | Summer) = P(Summer | Yes)* P(Yes) / P(Summer) = (0.33 x 0.64) /0.36 = 0.60

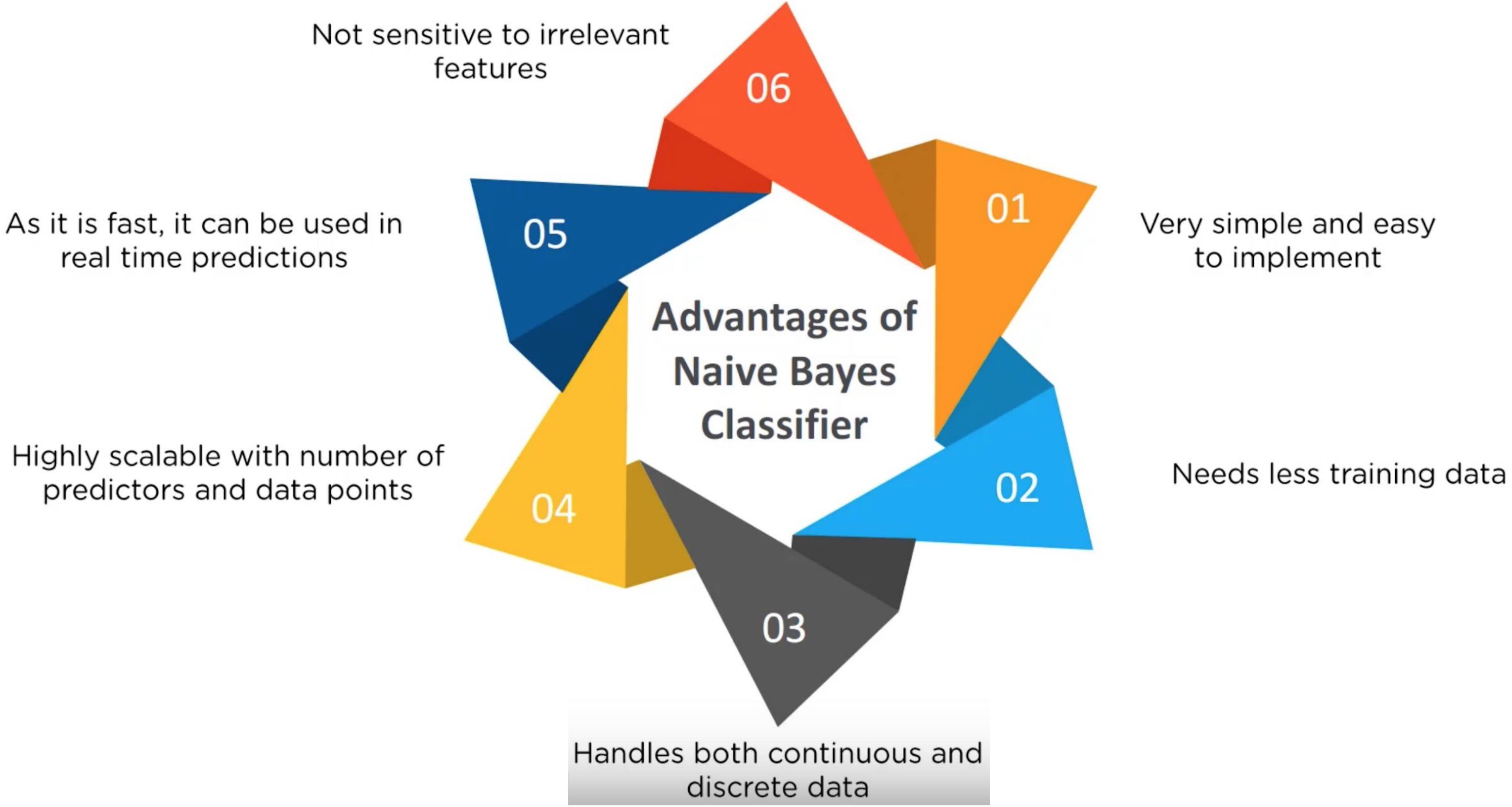
Let us use the likelihood table to predict whether to play football on (Season = Winter, Sunny = No , Windy = Yes)

$$P(c \mid x) = P(Play = Yes \mid Winter, Sunny = No, Windy = Yes)$$

$$= \frac{P(Winter \mid Yes) * P(Sunny = No \mid Yes) * P(Windy = Yes \mid Yes) * P(Yes)}{P(Winter) * P(Sunny = No) * P(Windy = Yes)}$$

$$= (2/9) * (6/9) * (6/9) * (9/14) \;/\; (5/14) * (7/14) * (8/14) = 0.6223$$

Since the probability is greater than 0.5, we should play football on that day.


Advantages of Naive Bayes Classifier
01
Very simple and easy to implement
02
Needs less training data
03
Handles both continuous and discrete data
04
Highly scalable with number of predictors and data points
05
As it is fast, it can be used in real time predictions
06
Not sensitive to irrelevant features