Introduction to Naïve Bayes

Course Overview

You are here...

Term	CDF	GCD	GCDAI	PGPDSAI
Term 1	Data Analytics with Python	Data Analytics with Python	Data Analytics with Python	Data Analytics with Python
Term 2	Data Visualization Techniques	Data Visualization Techniques	Data Visualization Techniques	Data Visualization Techniques
Term 3	EDA & Data Storytelling	EDA & Data Storytelling	EDA & Data Storytelling	EDA & Data Storytelling
		Minor Project	Minor Project	Minor Project
Term 4		Machine Learning Foundation	Machine Learning Foundation	Machine Learning Foundation
Term 5		Machine Learning Intermediate	Machine Learning Intermediate	Machine Learning Intermediate
Term 6		Machine Learning Advanced (Mandatory)	Machine Learning Advanced (Mandatory)	Machine Learning Advanced (Mandatory)
		Data Visualization with Tableau (Elective - I)	Data Visualization with Tableau (Elective - I)	Data Visualization with Tableau (Elective - I)
		Data Analytics with R (Elective - II)	Data Analytics with R (Elective - II)	Data Analytics with R (Elective - II)
		Capstone Project	Capstone Project	Capstone Project
Term 7		Bonus: Industrial ML (ML – 4 & 5)	Basics of AI, TensorFlow, and Keras	Basics of AI, TensorFlow, and Keras
Term 8			Deep Learning Foundation	Deep Learning Foundation
Term 9			NPL – I/CV – I	CV – I
Term 10			NLP – II/CV – II	NLP – I
			Capstone Project	Capstone Project
Term 11				CV – II
Term 12				NLP – II
				NLP – III + CV – III
				AutoVision & AutoNLP
				Building AI product

Term Context

- Decision Tree
- Random Forest
- Principal Component Analysis
- **Naïve Bayes Classifier** You are here…

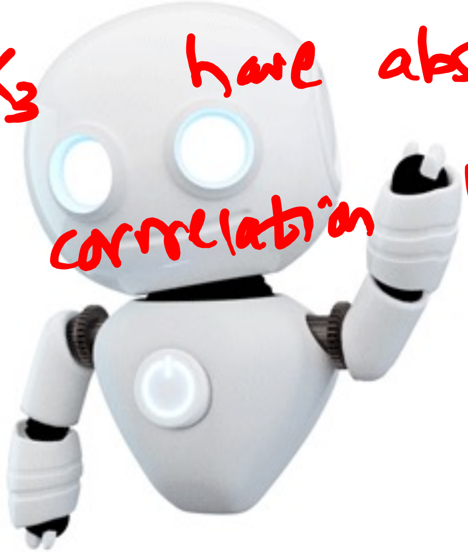
Agenda

1. **What is Naïve Bayes?** → Baye's Thm
Naya hai waha
2. Why it is called Naïve?
3. Bayes Theorem & Example
4. Naïve Bayes In Working
5. Application of Naïve Bayes
6. Pros/Cons of Naïve Bayes

What is Naïve Bayes?

- A family of simple "**probabilistic classifiers**“.
- Based on applying **Bayes' theorem** with strong (naïve) **independence** assumptions between the features.

$X_1 + X_2 + X_3 \sim y \rightarrow$ N.B. algo. will give fantastic results iff X_1, X_2 & X_3 have absolutely no correlation or marginal correlation b/w each other.



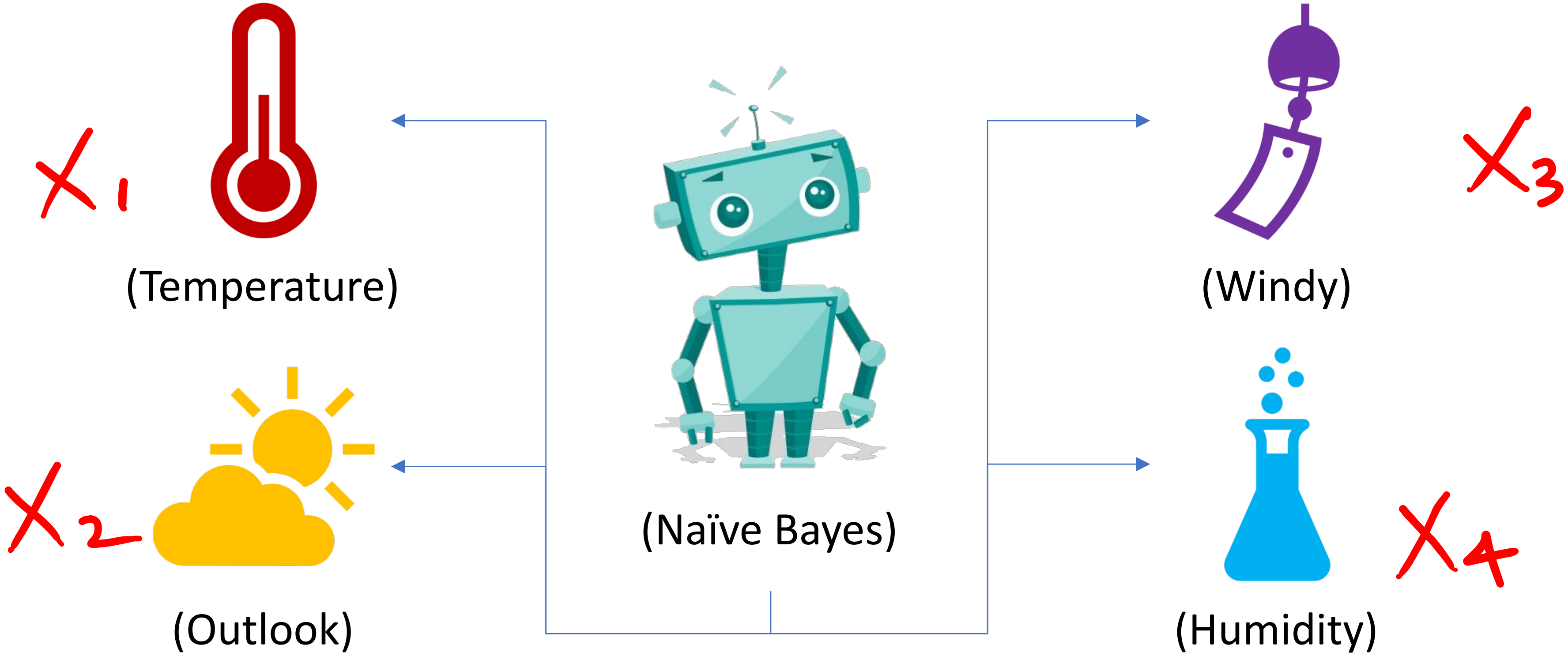
Agenda

1. What is Naïve Bayes?
2. **Why it is called Naïve?**
3. Bayes Theorem & Example
4. Naïve Bayes In Working
5. Application of Naïve Bayes
6. Pros/Cons of Naïve Bayes

INSAID

Why it is called Naïve?

- It assume all attributes to be conditionally **independent**.
- **Example:** Features describing whether it is a good day to play golf or not.



INSAID

Agenda

INSAID

Bayes Theorem & Example

- It was named after **Reverend Thomas Bayes**.
- It describes the **probability** of an **event** happening.
- It is based on **prior knowledge** of conditions that might be related to the event.

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

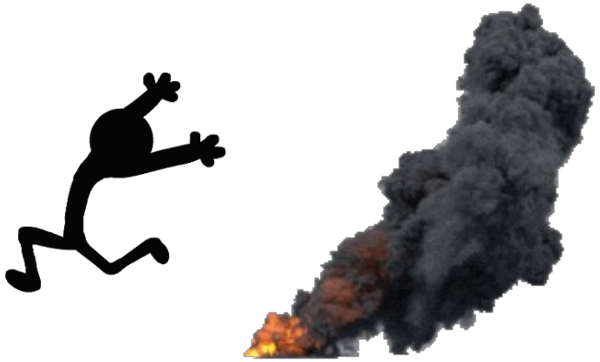
- A, B = Events
- $P(A|B) =$ Probability of A given B is true
- $P(A|B) =$ Probability of B given A is true
- $P(A), P(B) =$ Independent probabilities of A and B

Bayes Theorem & Example

P(Fire) P(Smoke)

- Let us say **dangerous fires** are rare **(1%)**, but **smoke** is fairly common **(10%)** due to barbecues,
- We add a condition that **90%** of **dangerous fires make smoke**.

P(Smoke | Fire)



- $P(\text{Fire})$ = how often there is fire
- $P(\text{Smoke})$ = how often we see smoke
- $P(\text{Fire}|\text{Smoke})$ = how often there is fire when we see smoke
- $P(\text{Smoke}|\text{Fire})$ = how often we see smoke when there is fire

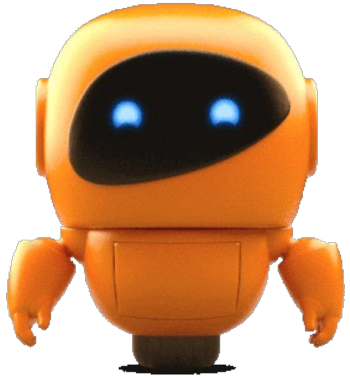
Bayes Theorem & Example

- We can then discover the probability of **dangerous Fire** when there is **Smoke**:

$$\mathrm{P(Fire|Smoke)} = \frac{\mathrm{P(Smoke|Fire) \cdot P(Fire)}}{\mathrm{P(Smoke)}}$$

$$\mathrm{P(Fire|Smoke)} = \frac{90\% \,.\, 1\%}{10\%}$$

$$\mathrm{P(Fire|Smoke)} = 9\%$$



- So it is still **worth checking** out for any **smoke** as a **precaution**.

Agenda

1. What is Naïve Bayes?
2. Why it is called Naïve?
3. Bayes Theorem & Example
4. **Naïve Bayes In Working**
5. Application of Naïve Bayes
6. Pros/Cons of Naïve Bayes

Naïve Bayes In Working

y

- Let's say we have the following data using which we want to **predict** whether to **play golf** or **not**.

x_1 x_2 x_3 x_4

#	Outlook	Temperature	Humidity	Windy	Play
01	Sunny	Hot	High	False	No
02	Sunny	Hot	High	True	No
03	Overcast	Hot	High	False	Yes
04	Rainy	Mild	High	False	Yes
05	Rainy	Cool	Normal	False	Yes
06	Rainy	Cool	Normal	True	No
07	Overcast	Cool	Normal	True	Yes
08	Sunny	Mild	High	False	No
09	Sunny	Cool	Normal	False	Yes
10	Rainy	Mild	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Rainy	Mild	High	True	No

$n=14$

Yes 9

No 5

Naïve Bayes In Working

14

#	Play
01	No
02	No
03	Yes
04	Yes
05	Yes
06	No
07	Yes
08	No
09	Yes
10	Yes
11	Yes
12	Yes
13	Yes
14	No

Total = 14

9 Yes

#	Play
01	Yes
02	Yes
03	Yes
04	Yes
05	Yes
06	Yes
07	Yes
08	Yes
09	Yes

Total Yes = 9

5 No

#	Play
01	No
02	No
03	No
04	No
05	No

Total No = 5

Naïve Bayes In Working

- First we need to identify the probability of **Play = Yes** and **Play = No**.

#	Play	P(Yes/No)
01	Yes 9	9 ÷ 14
02	No 5	5 ÷ 14
03	**Total** 14	100%

- Now we need to calculate **probability** of **each features** with respect to **Play Feature**.

X_1:

#	Outlook Calculation				
	Outlook	Yes	No	P(Yes)	P(No)
01	Sunny	2	3	2 ÷ 9	3 ÷ 5
02	Overcast	4	0	4 ÷ 9	0 ÷ 5
03	Rainy	3	2	3 ÷ 9	2 ÷ 5
04	**Total**	9	5	100%	100%

X_2:

#	Temperature Calculation				
	Temperature	Yes	No	P(Yes)	P(No)
01	Hot	2	2	2 ÷ 9	2 ÷ 5
02	Mild	4	2	4 ÷ 9	2 ÷ 5
03	Cool	3	1	3 ÷ 9	1 ÷ 5
04	**Total**	9	5	100%	100%

Naïve Bayes In Working

X_3 :

#	Humidity Calculation				
	Humidity	Yes	No	P(Yes)	P(No)
01	High	3	4	3 ÷ 9	4 ÷ 5
02	Normal	6	1	6 ÷ 9	1 ÷ 5
03	**Total**	9	5	100%	100%

X_4 :

#	Windy Calculation				
	Windy	Yes	No	P(Yes)	P(No)
01	False	6	2	6 ÷ 9	2 ÷ 5
02	True	3	3	3 ÷ 9	3 ÷ 5
03	**Total**	9	5	100%	100%

- We have finally **obtained** all the **values** that are required to a **predict** a **new value**.
- Let’s say we have a new instance as following:

Unseen Data →

#	Outlook	Temperature	Humidity	Windy	Play
01	Sunny	Cool	High	True	?

Naïve Bayes In Working

- Now we need to estimate the **probability** that we can **play game** or **not** using those **lookup tables**.

Lookup Table					
	Categories	Yes	No	P(Yes)	P(No)
Outlook	Sunny	2	3	2 ÷ 9	3 ÷ 5
	Overcast	4	0	4 ÷ 9	0 ÷ 5
	Rainy	3	2	3 ÷ 9	2 ÷ 5
	Total	9	5	100%	100%
Temperature	Hot	2	2	2 ÷ 9	3 ÷ 5
	Mild	4	2	4 ÷ 9	2 ÷ 5
	Cool	3	1	3 ÷ 9	1 ÷ 5
	Total	9	5	100%	100%
Humidity	High	3	4	3 ÷ 9	4 ÷ 5
	Normal	6	1	6 ÷ 9	1 ÷ 5
	Total	9	5	100%	100%
Windy	False	6	2	6 ÷ 9	2 ÷ 5
	True	3	3	3 ÷ 9	3 ÷ 5
	Total	9	5	100%	100%

Probability that we can play a game:

- P(Outlook = Sunny | Play = Yes) = 2 ÷ 9
- P(Temperature = Cool | Play = Yes) = 3 ÷ 9
- P(Humidity = High | Play = Yes) = 3 ÷ 9
- P(Windy = True | Play = Yes) = 3 ÷ 9
- P(Play = Yes) = 9 ÷ 14

Naïve Bayes In Working

- Now we need to estimate the **probability** that we can **play game** or **not** using those **lookup tables**.

Lookup Table					
	Categories	Yes	No	P(Yes)	P(No)
Outlook	Sunny	2	3	2 ÷ 9	3 ÷ 5
	Overcast	4	0	4 ÷ 9	0 ÷ 5
	Rainy	3	2	3 ÷ 9	2 ÷ 5
	Total	9	5	100%	100%
Temperature	Hot	2	2	2 ÷ 9	3 ÷ 5
	Mild	4	2	4 ÷ 9	2 ÷ 5
	Cool	3	1	3 ÷ 9	1 ÷ 5
	Total	9	5	100%	100%
Humidity	High	3	4	3 ÷ 9	4 ÷ 5
	Normal	6	1	6 ÷ 9	1 ÷ 5
	Total	9	5	100%	100%
Windy	False	6	2	6 ÷ 9	2 ÷ 5
	True	3	3	3 ÷ 9	3 ÷ 5
	Total	9	5	100%	100%

Probability that we cannot play a game:

- P(Outlook = Sunny | Play = No) = 3 ÷ 5
- P(Temperature = Cool | Play = No) = 1 ÷ 5
- P(Humidity = High | Play = No) = 4 ÷ 5
- P(Windy = True | Play = No) = 3 ÷ 5
- P(Play = No) = 5 ÷ 14

Naïve Bayes In Working

- Next we will multiply all the probability for both the cases (being Yes or No) individually.

$$P(X \mid Play = Yes) \,.\, P(Play = Yes)$$

$$(2 \div 9) * (3 \div 9) * (3 \div 9) * (3 \div 9) * (9 \div 14)$$

0.0053

$$P(X \mid Play = No) \,.\, P(Play = No)$$

$$(3 \div 5) * (1 \div 5) * (4 \div 5) * (3 \div 5) * (5 \div 14)$$

0.0206

- Finally we need to **divide both** of the **result** by **evidence** P(X) to normalize.
- The **evidence** for both the equation is **same**.
- We can find the value from the Total column of the lookup table.

Naïve Bayes In Working

Unseen

- P(X) = P(Outlook = Sunny) * P(Temperature = Cool) * P(Humidity = High) * P(Windy =True)
- P(X) = (5 ÷ 14) * (4 ÷ 14) * (7 ÷ 14) * (6 ÷ 14)
- P(X) = 0.02186
- Finally we will use this value to divide the probability of playing Yes and No.

P(Play = Yes | X) = 0.0053 ÷ 0.02186 = 0.2424

P(Play = No | X) = 0.0206 ÷ 0.02186 = 0.9421

0.9421 > 0.2424

- Given the final outcome **we cannot play Golf**.

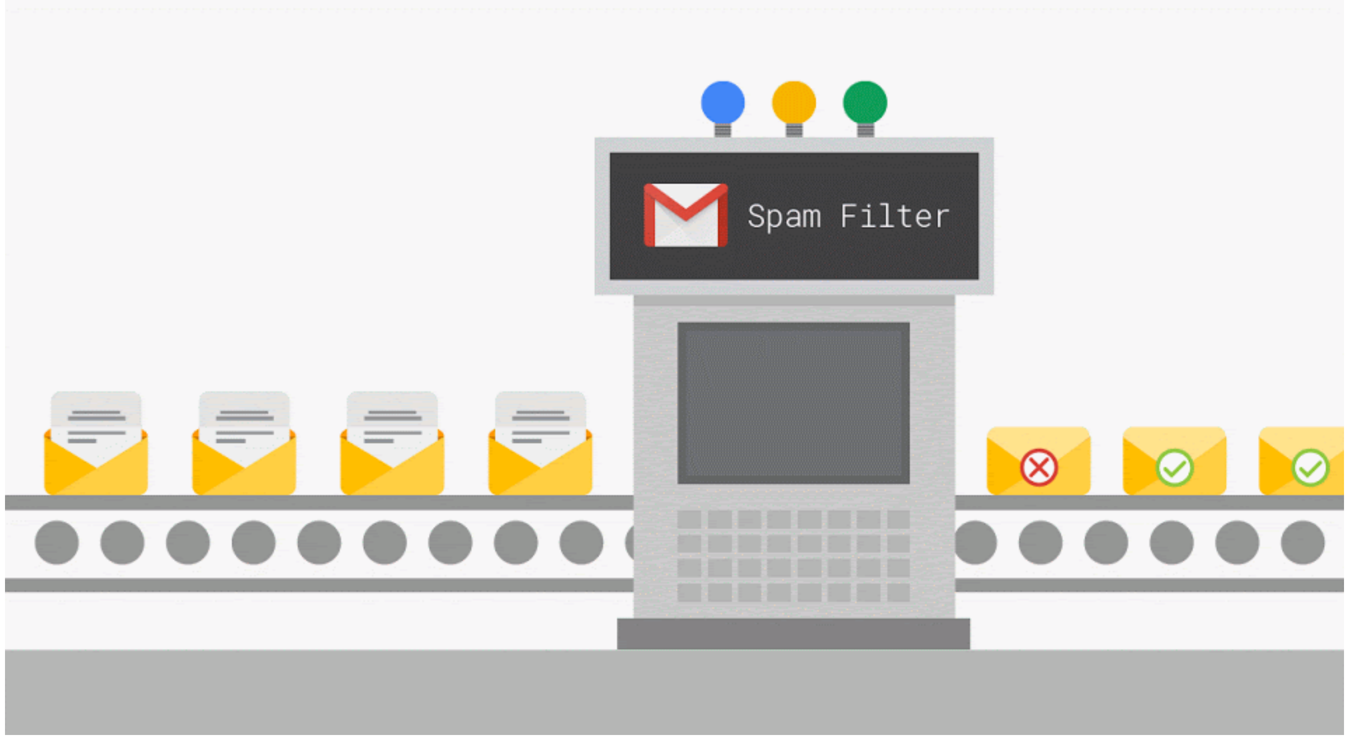
INSAID

Agenda

1. What is Naïve Bayes?
2. Why it is called Naïve?
3. Bayes Theorem & Example
4. Naïve Bayes In Working
5. **Application of Naïve Bayes**
6. Pros/Cons of Naïve Bayes

Spam Filtration

- Naïve bayes can also be used to **identify** whether the mail is **spam free** or **not**.



Weather Prediction

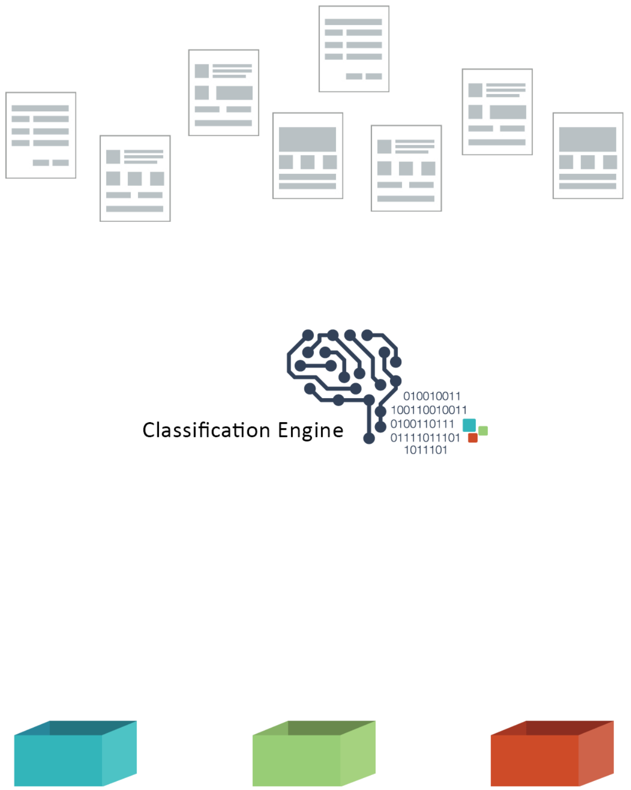
- We can **forecast** the **weather condition** of specific **region** based on the pre historic data .
- It can **save resources** and **prepare** for the **changes** coming forth.

Face Recognition

- It is considered as **automated** approach to verify the **identification** of the **human** being.
- It is done based on characteristics such as fingerprint, iris pattern, or face.

Text Classification (NLP)

- Naïve Bayes have a **higher success rate** in terms of **classifying text** as compared to other algorithms.



Agenda

1. What is Naïve Bayes?
2. Why it is called Naïve?
3. Bayes Theorem & Example
4. Naïve Bayes In Working
5. Application of Naïve Bayes
6. **Pros/Cons of Naïve Bayes**

Pros/Cons of Naïve Bayes

Pros

- It requires **less training data** and **handles** both **categorical** and **discrete** data.
- It is very **simple**, **fast** and **easy** to implement even in real time.
- Highly **scalable** with number of data points.
- Not **sensitive** to **irrelevant** features.

Cons

- **Assumption** of **independent** predictors, which is **impossible** in real life.
- Categories present in the test dataset, which was not observed in training (Also know as **Zero Frequency**).
- To handle zero frequency problem **smoothing** techniques are used which we will cover later on.

Thank You