Text Classification

Mohamed Elhaj-Abdou

Text Classification

Recap on what is classification

Examples of Text classification

Mohamed Elhaj-Abdou

What is Classification

Mohamed Elhaj-Abdou

What is Text Classification

Mohamed Elhaj-Abdou

The data is Text such as

Tweet Or Post

Positive

Negative

Article

Sport

Politics

tech

Mohamed Elhaj-Abdou

Lets take an example

Text	Class
I love mom	Positive
I love dad	Positive
I hate lairs	Negative
I hate ice Creem in winter	Negative

It actually searching on a words that related to positive or negative such as

How we can represents this kind of data mathematically

We need to find what is the type of the input data in form of text

Structured or un-structured?

Mohamed Elhaj-Abdou

The text data is unstructured data

Why

The text's may not be have te same lengths

Doesn't have any pre-defined model
Its generated by different users

Not organized, you may find text related to sport and others

You may find different type of data such as text or dates or numbers→ this leads difficulties to model these data with the standard programs

Mohamed Elhaj-Abdou

Solutions to deal with unstructured data

Finding patterns such as Regular expressions

Text tagging

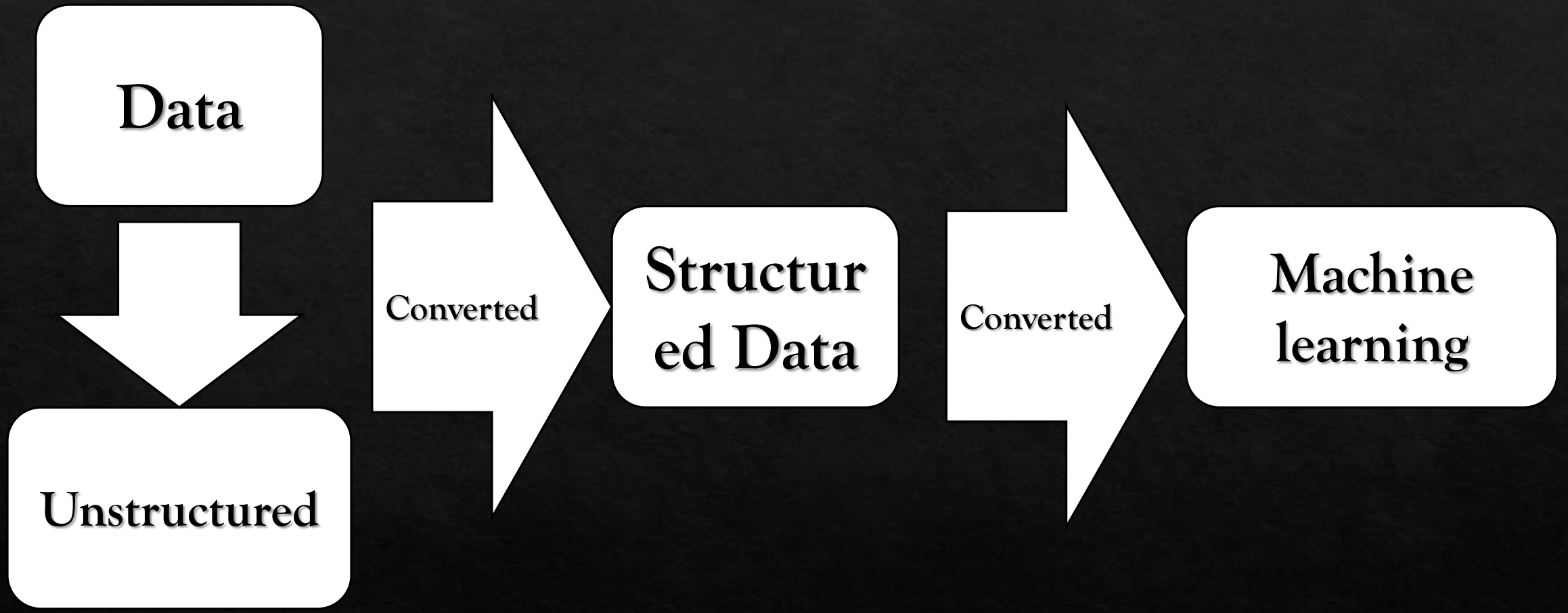
Text steaming

And many more...

Is machine learning model can deal with unstructured data ?

No

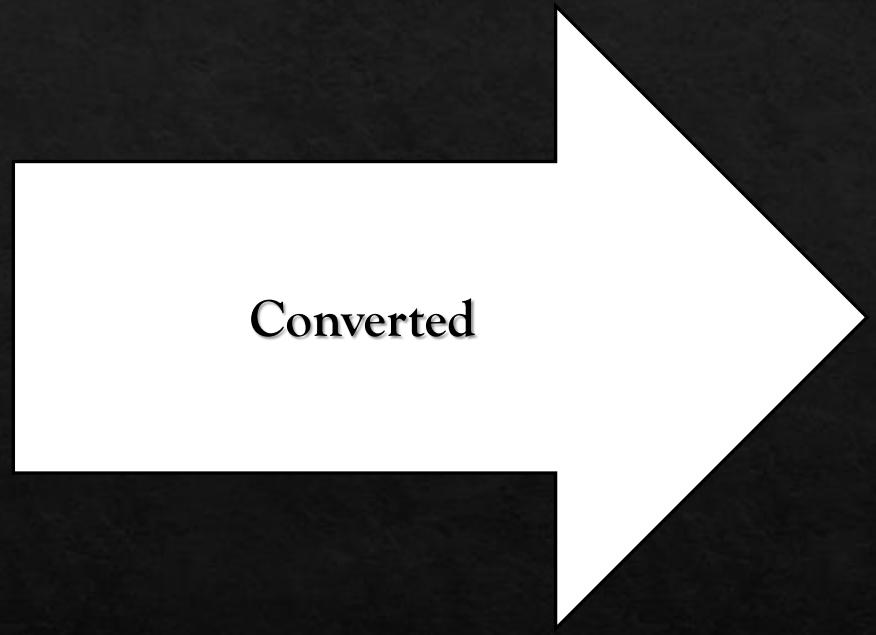
Mohamed Elhaj-Abdou



Mohamed Elhaj-Abdou

Unstructured

Text	Class
I love mom	Positive
I love dad	Positive
I hate lairs	Negative
I hate iceCreem	Negative



Structured

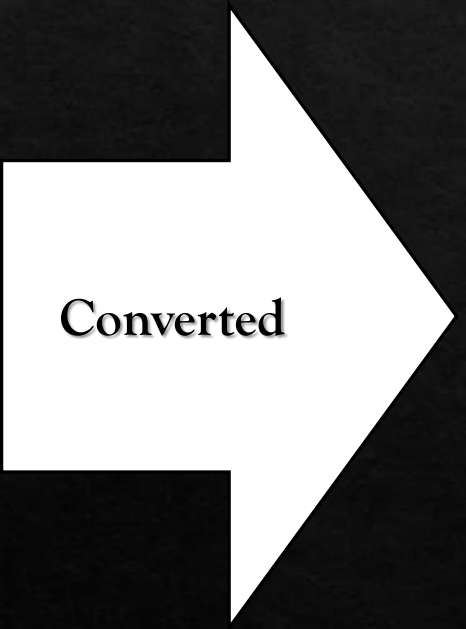
I love mom

I love dad

I hate lairs

I hate ice Creem

Structured



Number of sentences	I	Love	mom	lairs	dad	hate	iceCreem
1	1	1	1	0	0	0	0
2	1	1	0	0	1	0	0
3	1	0	0	1	0	1	0
4	1	0	0	0	0	1	1

This method named count vectorizer

Mohamed Elhaj-Abdou

count vectorizer

Method in NLP to convert the text into digital information

It convert each sentence into vector

Count each word in each sentence

Mohamed Elhaj-Abdou

Inserting the count in vector

Each vector represents each sentences with numbers represents the occurrences (count)

Each row → vector→ each vector → sentence

Mohamed Elhaj-Abdou