

# There are various approaches for converting text into vector

Label  
Encoding

One Hot  
Encoding

Bag Of Words

TF-IDF

Word  
Embeddings



# There are various approaches of converting text into vector

Label  
Encoding

One Hot  
Encoding

Bag Of Words

TF-IDF

Word  
Embeddings



**Urgent Business Assistance, There is 55 million \$ in your account, claim it hurry up**



Urgent Business Assistance, There is 55 million \$ in your account, claim it hurry up

**Text Representation**

**Feature Engineering**

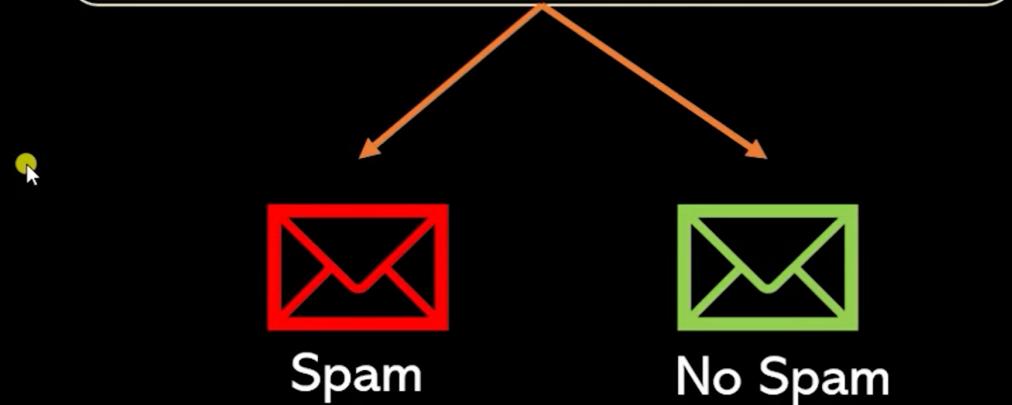
[ 3, 0, 1, 0, 0, 0, 6, ..., 0 ]

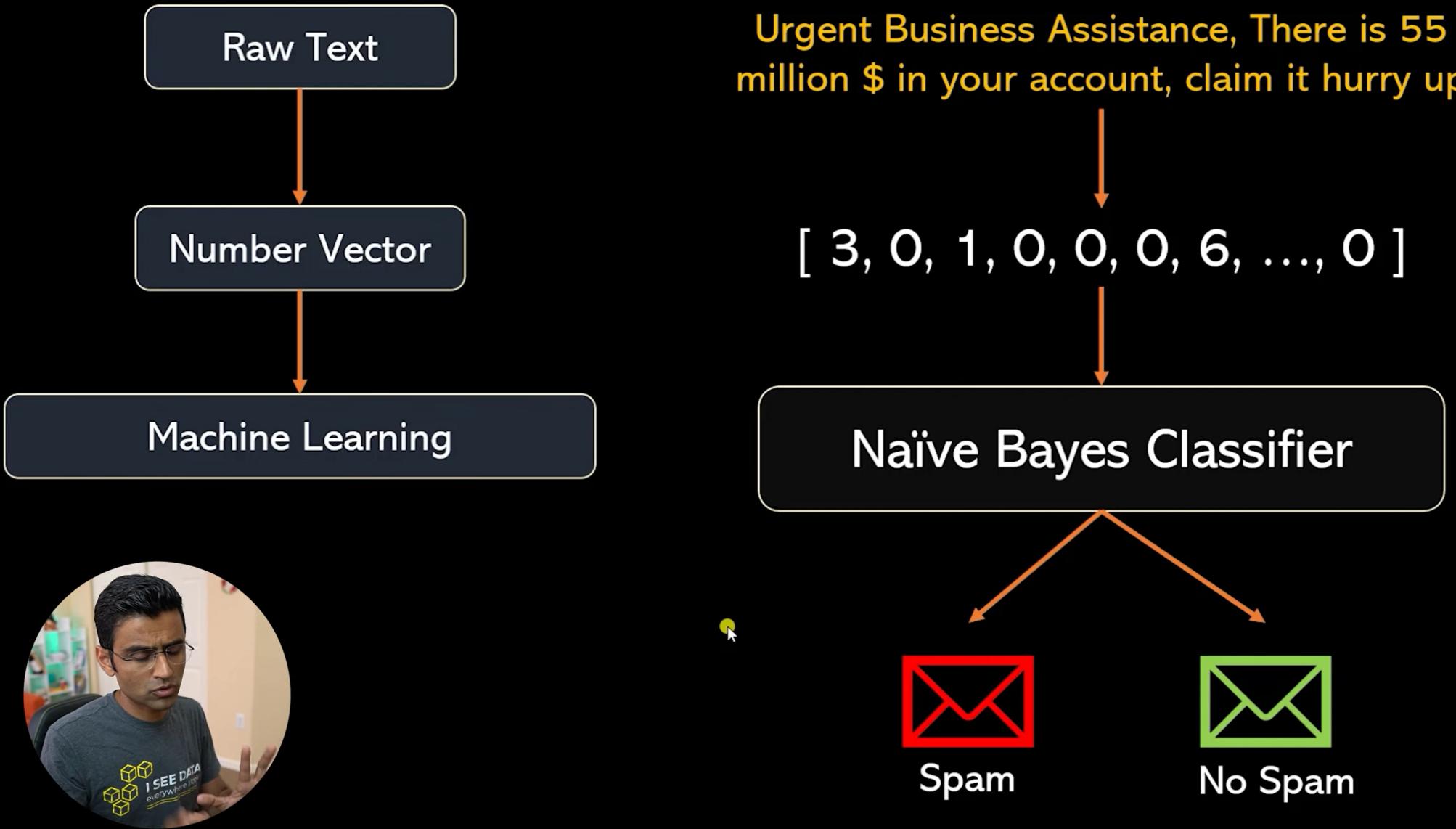


Urgent Business Assistance, There is 55 million \$ in your account, claim it hurry up

[ 3, 0, 1, 0, 0, 0, 6, ..., 0 ]

Naïve Bayes Classifier







Hey Pranav, can you add auto play next video feature on [codebasics.io](http://codebasics.io)



Urgent Business Assistance, There is 55 million \$ in your account, claim it hurry up



Most of the participants have just heard about the topic Data analytics and are not having much knowledge on it.





Hey Pranav, can you add auto play next video feature on [codebasics.io](#)



Urgent Business Assistance, There is 55 million \$ in your account, claim it hurry up



Most of the participants have just heard about the topic Data analytics and are not having much knowledge on it.



add, auto, and, assistance, are,

business, can, claim, [codebsics.io](#),

data, 55, hurry,

...

urgent

## VOCABULARY

Visit [codebasics.io](#) for my practical, affordable video courses



Hey Pranav, can you add auto play  
next video feature on codebasics.io

add, auto, and, assistance, are,  
business, can, claim, codebsics.io,  
data, 55, hey,  
...  
urgent

## VOCABULARY



Hey Pranav, can you add auto play  
next video feature on codebasics.io



1 2 3 4 5  
add, auto, and, assistance, are,  
6 7 8 9  
business, can, claim, codebsics.io,  
10 11 12  
data, 55, hey,  
...  
urgent  
2000

## VOCABULARY

Hey Pranav, can you add auto play  
next video feature on codebasics.io



```
[ 12 1087 7 1870 1 2 ... 9 ]
```

1 2 3 4 5  
add, auto, and, assistance, are,  
6 7 8 9  
business, can, claim, codebsics.io,  
10 11 12  
data, 55, hey,  
...  
urgent  
2000

## VOCABULARY



Hey Pranav, can you add auto play  
next video feature on codebasics.io



```
[ 12 1087 7 1870 1 2 ... 9 ]
```

1 2 3 4 5  
add, auto, and, assistance, are,  
6 7 8 9  
business, can, claim, codebsics.io,  
10 11 12  
data, 55, hey,  
...  
urgent  
2000

**VOCABULARY**



Hey Pranav, can you add auto play  
next video feature on codebasics.io



```
[ 12 1087 7 1870 1 2 ... 9 ]
```



1 2 3 4 5

add, auto, and, assistance, are,

6 7 8 9

business, can, claim, codebsics.io,

10 11 12

data, 55, hey,

...

urgent

2000

**VOCABULARY**

Hey Pranav, can you add auto play  
next video feature on codebasics.io



[ 12 1087 7 1870 1 2 ... 9 ]



Label  
Encoding



1 2 3 4 5  
add, auto, and, assistance, are,  
6 7 8 9  
business, can, claim, codebsics.io,  
10 11 12  
data, 55, hey,  
...  
urgent  
2000

## VOCABULARY

add, auto, and, assistance, are, business, can, ... urgent



add, auto, and, assistance, are, business, can, ... urgent

Hey

Pranav

can

you

add

auto

...



add, auto, and, assistance, are, business, can, ... urgent

Hey	[	0	0	0	0	0	0	0	...	0	]
Pranav	[	0	0	0	0	0	0	0	...	0	]
can	[	0	0	0	0	0	0	1		0	]
you											
add											
auto											
...											



add, auto, and, assistance, are, business, can, ... urgent

Hey [ 0 0 0 0 0 0 0 ... 0 ]

Pranav [ 0 0 0 0 0 0 ... 0 ]

can [ 0 0 0 0 0 0 1 ... 0 ]

you

add

auto

...



add, auto, and, assistance, are, business, can, ... urgent

Hey	[	0	0	0	0	0	0	0	...	0	]
Pranav	[	0	0	0	0	0	0	0	...	0	]
can	[	0	0	0	0	0	0	1		0	]
you	[	0	0	0	0	0	0	0		0	]
add	[	1	0	0	0	0	0	0		0	]
auto	[	0	1	0	0	0	0	0		0	]

...



Hey	[	0	0	0	0	0	0	0	...	0	]
Pranav	[	0	0	0	0	0	0	0	...	0	]
can	[	0	0	0	0	0	0	0	1	0	]
you	[	0	0	0	0	0	0	0	0	0	]
add	[	1	0	0	0	0	0	0	0	0	]
auto	[	0	1	0	0	0	0	0	0	0	]

One Hot  
Encoding



Hey Pranav, can you add auto play  
next video feature on codebasics.io

[ 12 1087 7 1870 1 2 ... 9 ]



Hey	[	0	0	0	0	0	0	0	...	0 ]
Pranav	[	0	0	0	0	0	0	0	...	0 ]
can	[	0	0	0	0	0	0	1		0 ]
you	[	0	0	0	0	0	0	0		0 ]
add	[	1	0	0	0	0	0	0		0 ]
auto	[	0	1	0	0	0	0	0		0 ]



Hey Pranav, can you add auto play  
next video feature on codebasics.io

[ 12 1087 7 1870 1 2 ... 9 ]



Hey	[	0	0	0	0	0	0	0	...	0 ]
Pranav	[	0	0	0	0	0	0	0	...	0 ]
can	[	0	0	0	0	0	0	1		0 ]
you	[	0	0	0	0	0	0	0		0 ]
add	[	1	0	0	0	0	0	0		0 ]
auto	[	0	1	0	0	0	0	0		0 ]

Label  
Encoding

One Hot  
Encoding



# Disadvantages of label encoding and one hot encoding



add, auto, and, assistance, need, I, immediately, help, urgent

I	[	0	0	0	0	0	1	0	0	0	]
need	[	0	0	0	0	1	0	0	0	0	]
help	[	0	0	0	0	0	0	0	1	0	]



add, auto, and, assistance, need, I, immediately, help, urgent

I	[	0	0	0	0	0	1	0	0	0	]
need	[	0	0	0	0	1	0	0	0	0	]
help	[	0	0	0	0	0	0	0	1	0	]

I	[	0	0	0	0	0	1	0	0	0	]
need	[	0	0	0	0	1	0	0	0	0	]
assistance	[	0	0	0	1	0	0	0	0	0	]



	add,	auto,	and,	assistance,	need,	I,	immediately,	help,	urgent
I	[	0	0	0	0	0	1	0	0 ]
need	[	0	0	0	0	1	0	0	0 ]
help	[	0	0	0	0	0	0	1	0 ]
I	[	0	0	0	0	0	1	0	0 ]
need	[	0	0	0	0	1	0	0	0 ]
assistance	[	0	0	0	1	0	0	0	0 ]



	add,	auto,	and,	assistance,	need,	I,	immediately,	help,	urgent	
I	[	0	0	0	0	0	1	0	0	]
need	[	0	0	0	0	1	0	0	0	]
help	[	0	0	0	0	0	0	0	1	0
I	[	0	0	0	0	0	1	0	0	]
need	[	0	0	0	0	1	0	0	0	]
assistance	[	0	0	0	1	0	0	0	0	]



add, auto, and, assistance, need, I, immediately, help, urgent

I	[	0	0	0	0	0	1	0	0	0	]
---	---	---	---	---	---	---	---	---	---	---	---

need	[	0	0	0	0	1	0	0	0	0	]
------	---	---	---	---	---	---	---	---	---	---	---

help	[	0	0	0	0	0	0	0	1	0	]
------	---	---	---	---	---	---	---	---	---	---	---

I	[	0	0	0	0	0	1	0	0	0	]
---	---	---	---	---	---	---	---	---	---	---	---

need	[	0	0	0	0	1	0	0	0	0	]
------	---	---	---	---	---	---	---	---	---	---	---

assistance	[	0	0	0	1	0	0	0	0	0	]
------------	---	---	---	---	---	---	---	---	---	---	---

Similar words do not have similar representation



add, auto, and, assistance, need, I, immediately, ... urgent (100 k words)

I

[ 0 0 0 0 0 1 0 ... 0 ] ← vector size=100k



add, auto, and, assistance, need, I, immediately, ... urgent (100 k words)

I [ 0 0 0 0 0 1 0 ... 0 ] ← vector size=100k



Consumes too much memory & compute resources



Visit [codebasics.io](https://codebasics.io) for my practical, affordable video courses

add, auto, and, assistance, need, I, immediately, help, urgent

I	[	0	0	0	0	0	1	0	0	0	]
need	[	0	0	0	0	1	0	0	0	0	]
help	[	0	0	0	0	0	0	0	1	0	]

bahubali



add, auto, and, assistance, need, I, immediately, help, urgent

I	[	0	0	0	0	0	1	0	0	0	]
need	[	0	0	0	0	1	0	0	0	0	]
help	[	0	0	0	0	0	0	0	1	0	]



add, auto, and, assistance, need, I, immediately, help, urgent

I	[	0	0	0	0	0	1	0	0	0	]
need	[	0	0	0	0	1	0	0	0	0	]
help	[	0	0	0	0	0	0	0	1	0	]

bahubali



add, auto, and, assistance, need, I, immediately, help, urgent

I [ 0 0 0 0 0 0 1 0 0 0 0 ]

need [ 0 0 0 0 1 0 0 0 0 0 ]

help [ 0 0 0 0 0 0 0 0 1 0 ]

bahubali

???



add, auto, and, assistance, need, I, immediately, help, urgent

I	[	0	0	0	0	0	1	0	0	0	]
need	[	0	0	0	0	1	0	0	0	0	]
help	[	0	0	0	0	0	0	0	1	0	]
bahubali	???										

Out Of Vocabulary (OOV) problem





I need banana

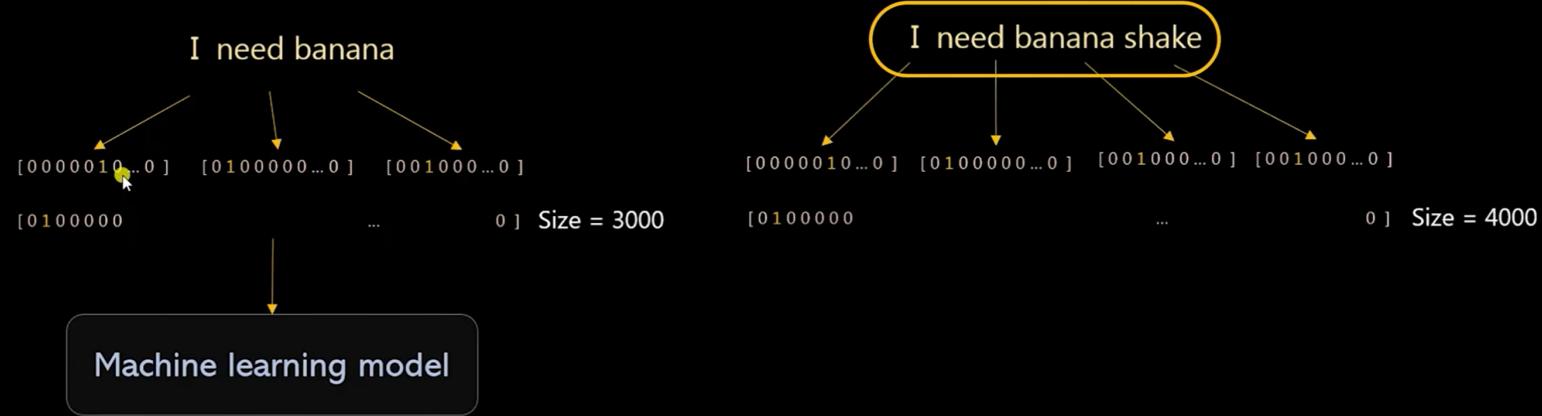


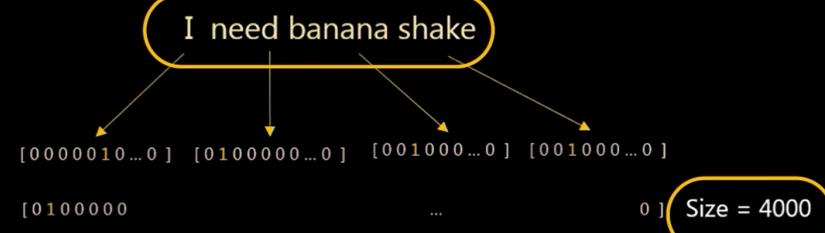
I need banana



Machine learning model







Machine learning model



I need banana

[00000010...0] [0100000...0] [001000...0]  
[0100000] ... 0 ] Size = 3000

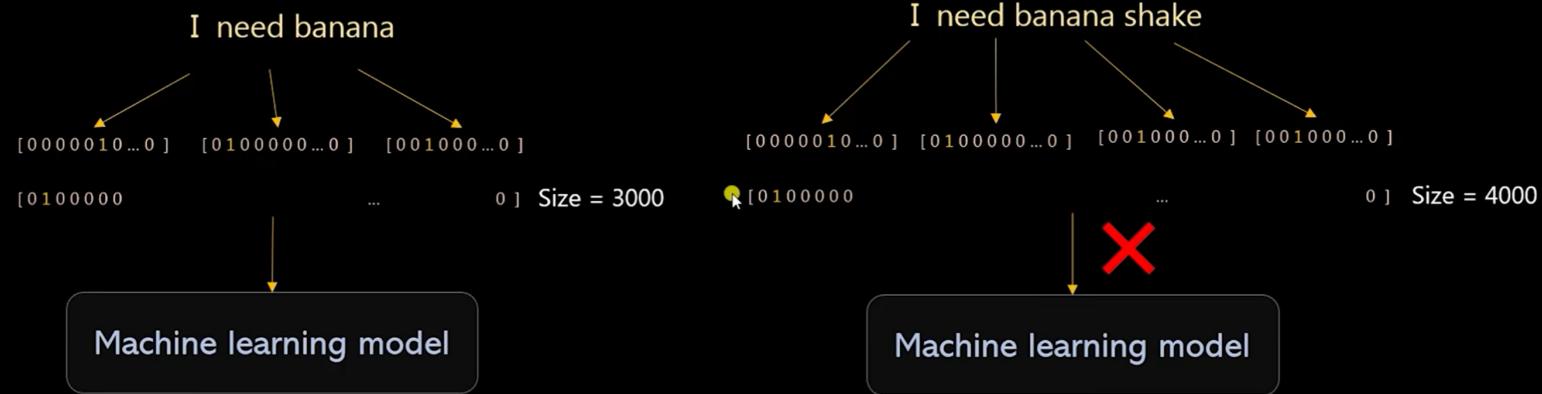
Machine learning model

I need banana shake

[00000010...0] [0100000...0] [001000...0] [001000...0]  
[0100000] ... 0 ] Size = 4000

Machine learning model





No fixed length representation

