



Natural Language Processing

Data Science

DETECTING SPAM (id=0)

From: Martha Washington <marthafirst@hotmail.com>

Subject: Lunch

Date: January 25, 1794 9:30:58 AM EST

To: George Washington <president@whitehouse.gov>

Looking forward to lunch at noon today. I reserved my MealPal at Nerds and Nibblers. Please don't forget to book yours. You know how much I hate standing in line.

xoxo Martha 🐱

DETECTING SPAM (id=1)

From: Aaron Burr <aabear@aol.com>

Subject: Quality meds, low price, quick delivery!

Date: January 05, 2004 9:05:01 AM EST

To: George Washington <gwashington@state.gov>

Not dead yet? ...ed!!!

GUARANTEED BEST PRICE VIAGRA. WE HAVE QUICKSILVER TOO!!!

CLICK HERE FOR VIAGRA! ORDER NOW!

<http://legit-pharmacy.ru/viagra-deals?ref=aabear>

[SPAM]

DETECTING SPAM (id=2)

From: George Washington <president@whitehouse.gov>

Subject: Need Viagra? We can help.

Date: January 05, 1704 0:05:00 AM EST

To: Martha Washington <martha@whitehouse.gov>

Not dead yet? We are not dead!!!

GUARANTEED BEST PRICE VIAGRA. WE HAVE QUICKSILVER TOO!!!

CLICK HERE FOR VIAGRA! ORDER NOW!

<http://legit-pharmacy.ru/viagra-deals?ref=aabear>

[SPAM]

DETECTING SPAM (id=3)

From: Martha Washington <marthafirst@hotmail.com>

Subject: No free lunch? Try cheap Viagra!

Date: January 05, 2004 9:05:00 AM EST

To: Alexander <alexander@duke.edu>

Not dead yet? ...ed!!!

GUARANTEED BEST PRICE VIAGRA. WE HAVE QUICKSILVER TOO!!!

CLICK HERE FOR VIAGRA! ORDER NOW!

<http://legit-pharmacy.ru/viagra-deals?ref=aabear>

[SPAM]

DETECTING SPAM (id=4)

From: George Washington <president@whitehouse.gov>
Subject: Re: Lunch
Date: January 25, 1794 10:12:58 AM EST
To: Martha Washington <marthafirst@hotmail.com>

OMG Nerds and Nibblers is my favorite lunch spot!
MealPal order confirmed! So excited for lunch. So tired
of these meetings! You're the best xoxo Geo 🥰

p.s. I got a weird email today. What's "Viagra"?



BAG OF WORDS

```
vocabulary = [ REDACTED ]
```

id	word_one	word_two	is_spam
0			False
1			True
2			True
3			True
4			False



CREATING DOCUMENT VECTORS

```
vocabulary = ['lunch', 'viagra']
```

id	lunch	viagra	is_spam
0	2	0	False
1	0	3	True
2	0	4	True
3	1	4	True
4	3	1	False



PLOTTING DOCUMENT VECTORS

```
vocabulary = ['lunch', 'viagra']
```

id	lunch	viagra	is_spam
0	2	0	False
1	0	3	True
2	0	4	True
3	1	4	True
4	3	1	False

