Attribute Detection

$$4 * 3 = 4 * n \quad * \quad n * 3$$
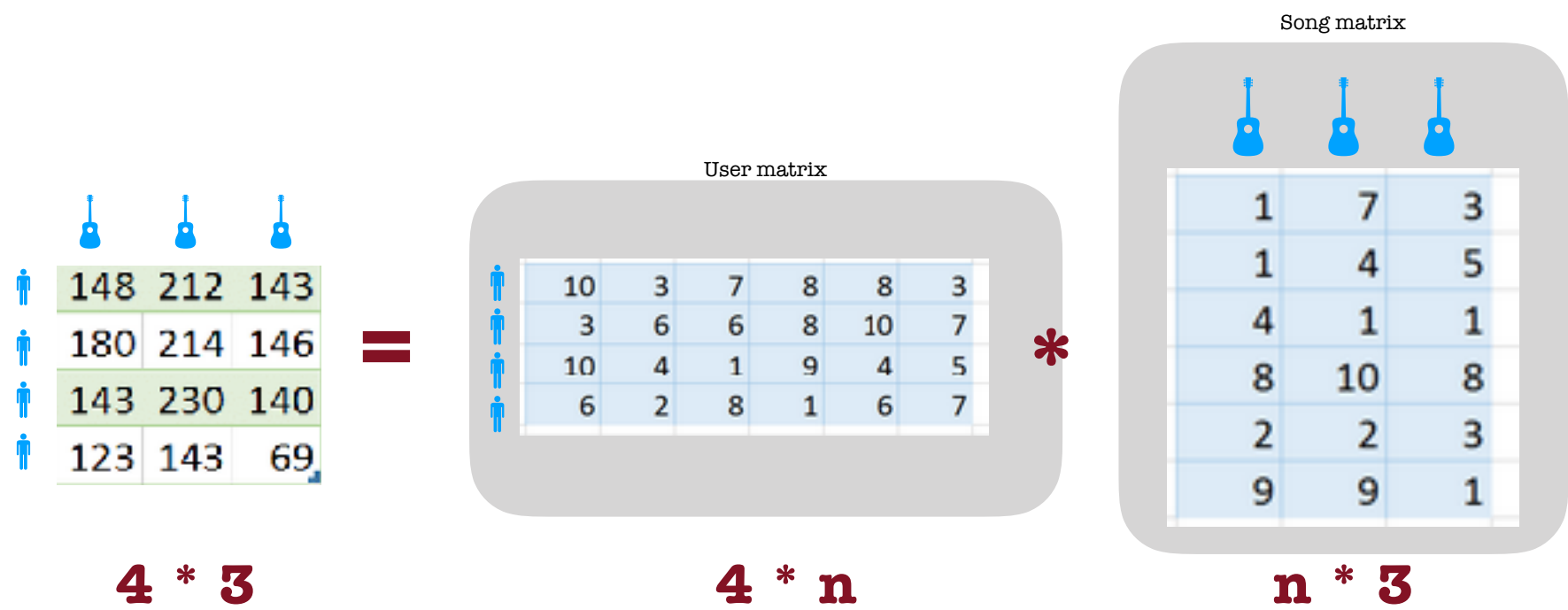
```
corpus = ["phone camera phone battery",
          "battery and camera of the phone",
          "cloth texture",
          "style and texture cloth"]
```

	phone	camera	battery	and	of	the	cloth	texture	style
doc_0	0.000000	0.408248	0.408248	0.000000	0.00000	0.816497	0.000000	0.000000	0.00000
doc_1	0.372225	0.372225	0.372225	0.000000	0.47212	0.372225	0.000000	0.000000	0.47212
doc_2	0.000000	0.000000	0.000000	0.707107	0.00000	0.000000	0.000000	0.707107	0.00000
doc_3	0.465809	0.000000	0.000000	0.465809	0.00000	0.000000	0.590819	0.465809	0.00000

Document matrix

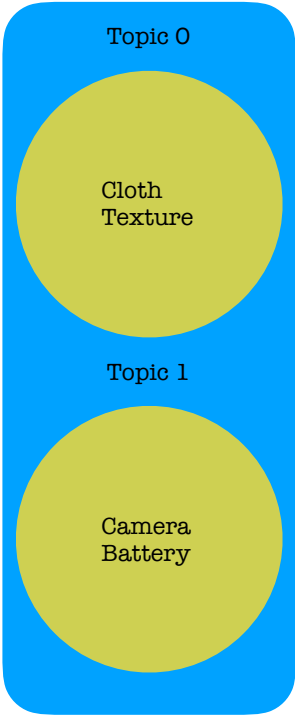|	0	1
doc_0	0.000000	0.878597
doc_1	0.056260	0.840770
doc_2	0.691503	0.000000
doc_3	0.707012	0.025084

- What we can infer from this?

If we just know the number of categories/topics present in our corpus we can find out which document belongs to which topic or which category

Topic 0: Doc 2, Doc 3
Topic 1: Doc 0, Doc 1
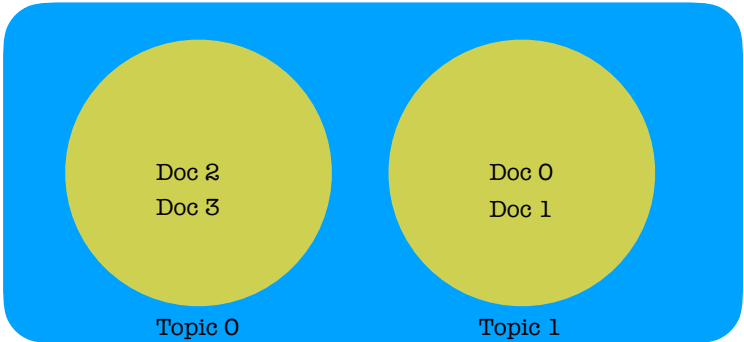
Word matrix

	0	1
phone	0.369680	0.183829
camera	0.000489	0.455349
battery	0.000489	0.455349
and	0.826328	0.000000
of	0.028011	0.256675
the	0.000000	0.707207
cloth	0.431368	0.000000
texture	0.826328	0.000000
style	0.028011	0.256675

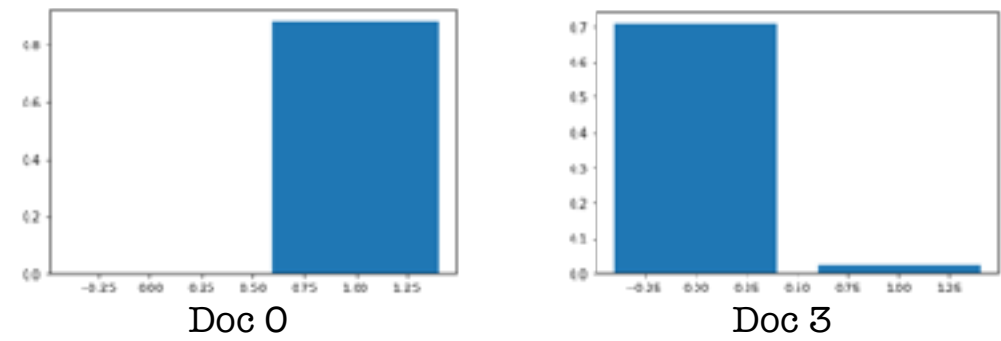Topic 0: Cloth Texture
Topic 1: Camera Battery

- What we can infer from this?

If we just know the number of categories/topics present in our corpus we can find out which word belongs to which topic or which category

Document matrix

	Topic 0	Topic 1
doc_0	0.000000	0.878597
doc_1	0.056260	0.840770
doc_2	0.691503	0.000000
doc_3	0.707012	0.025084

Document matrix represent the document distribution over topic

Doc 0

Doc 3

Word matrix

	Topic 0	Topic 1
phone	0.369680	0.183829
camera	0.000489	0.455349
battery	0.000489	0.455349
and	0.826328	0.000000
of	0.028011	0.256675
the	0.000000	0.707207
cloth	0.431368	0.000000
texture	0.826328	0.000000
style	0.028011	0.256675

Word matrix represent word distribution over topics
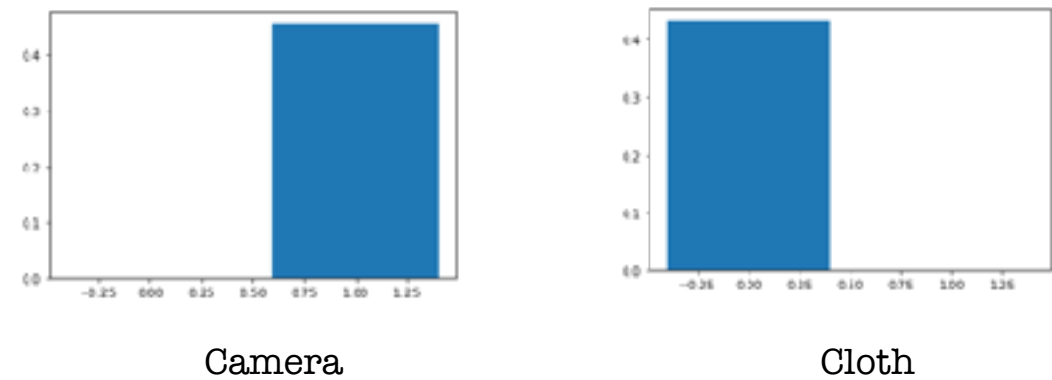
Camera

Cloth

Given a new document how will you find to what topic it belongs to?

New doc: cloth texture

TF-IDF representation

	phone	camera	battery	and	of	the	cloth	texture	style
new_doc	0.0	0.0	0.0	0.707107	0.0	0.0	0.0	0.707107	0.0

Find value of new_doc w.r.t Topic 0:

Word matrix

	Topic 0	Topic 1
phone	0.369680	0.183829
camera	0.000489	0.455349
battery	0.000489	0.455349
and	0.826328	0.000000
of	0.028011	0.256675
the	0.000000	0.707207
cloth	0.431368	0.000000
texture	0.826328	0.000000
style	0.028011	0.256675

Tf-IDF of new doc

	new_do
phone	0.000000
camer	0.000000
batter	0.000000
and	0.707107
of	0.000000
the	0.000000
cloth	0.000000
textur	0.707107
style	0.000000

```
= (0.369 * 0) +
(0.0004 * 0) +
So on
```

=0.69150341

Find value of new_doc w.r.t Topic 1:

Word matrix

	Topic 0	Topic 1
phone	0.369680	0.183829
camera	0.000489	0.455349
battery	0.000489	0.455349
and	0.826328	0.000000
of	0.028011	0.256675
the	0.000000	0.707207
cloth	0.431368	0.000000
texture	0.826328	0.000000
style	0.028011	0.256675

Tf-IDF of new doc

|	new_do
phone	0.000000
camer	0.000000
batter	0.000000
and	0.707107
of	0.000000
the	0.000000
cloth	0.000000
textur	0.707107
style	0.000000

= 0

Given:

```
corpus = ["phone camera phone battery",
          "battery and camera of the phone",
          "cloth texture",
          "style and texture cloth"]
```

	pho	camer	batter	and	of	the	cloth	textur	style
doc_0	0.000	0.4082	0.4082	0.0000	0.0000	0.8164	0.0000	0.0000	0.0000
doc_1	0.37	0.3722	0.3722	0.0000	0.472	0.3722	0.0000	0.0000	0.472
doc_2	0.000	0.0000	0.0000	0.7071	0.0000	0.0000	0.0000	0.7071	0.0000
doc_3	0.46	0.0000	0.0000	0.4658	0.0000	0.0000	0.5908	0.4658	0.0000

Document matrix

	Topic 0	Topic 1
doc_0	0.00000	0.878597
doc_1	0.05626	0.840770
doc_2	0.69150	0.000000
doc_3	0.70701	0.025084
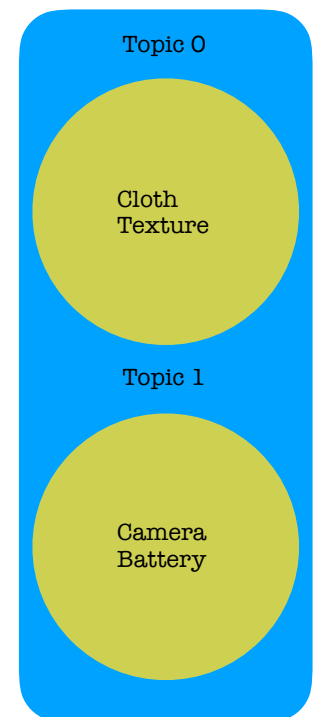
Word matrix

	Topic 0	Topic 1
phone	0.369680	0.183829
camera	0.000489	0.455349
battery	0.000489	0.455349
and	0.826328	0.000000
of	0.028011	0.256675
the	0.000000	0.707207
cloth	0.431368	0.000000
texture	0.826328	0.000000
style	0.028011	0.256675

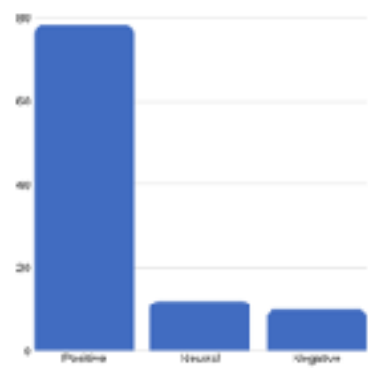Topic 0

Cloth Texture

Topic 1

Camera Battery

Case study 1:

How is the camera of iPhone X?

Product popularity

Product sentiment

Attributes:

- ☐ Camera
- ☐ Display
- ☐ Battery
- ☐ Screen
- ☐ Durability
- ☐ Quality
- ☐ Charging
- ☐ Speaker

Attribute popularity

Camera of this phone is very good and I fell in love with this phone...

URL: https://www.flipkart.com › ... › Apple iPhone 6s (Gold, 32 GB)
Summary: Camera is very good
● positive

Camera of this phone is very good and I fell in love with this phone...

URL: https://www.amazon.com › ... › Apple iPhone 6s (Gold, 32 GB)
Summary: Camera is very good
● positive

Camera of this phone is very good and I fell in love with this phone...
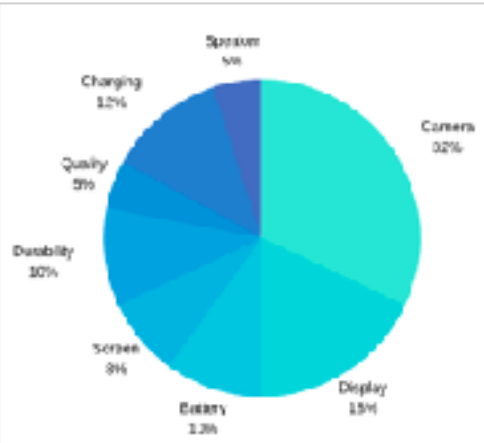
URL: https://www.snapdeal.com › ... › Apple iPhone 6s (Gold, 32 GB)
Summary: Camera is very good
● negative

Camera of this phone is very good and I fell in love with this phone...

URL: https://www.amazon.com › ... › Apple iPhone 6s (Gold, 32 GB)
Summary: Camera is very good
● neutral

Review 1 : Iphone x has very good camera. I bought this one in a local store near to my home. It has very good features.

Review 2 : Samsung s9 is one of the best phone. It has a good sound quality. The build is also very good.

Review 3 : Hello guys, I purchased this awesome shirt on Amazon with such a good price. I had to sell my kidney in order to buy this. But I feel it is worth selling my kindey after wearing it.

Review 4 : This shirt has very good material. Texture and style is what makes this brand awesome.

Review 1 : Iphone x has very good camera. I bought this one in a local store near to my home. It has very good features.

Review 2 : Samsung s9 is one of the best phone. It has a good sound quality. The build is also very good.

Review 3 : Hello guys, I purchased this awesome shirt on Amazon with such a good price. I had to sell my kidney in order to buy this. But I feel it is worth selling my kindey after wearing it.
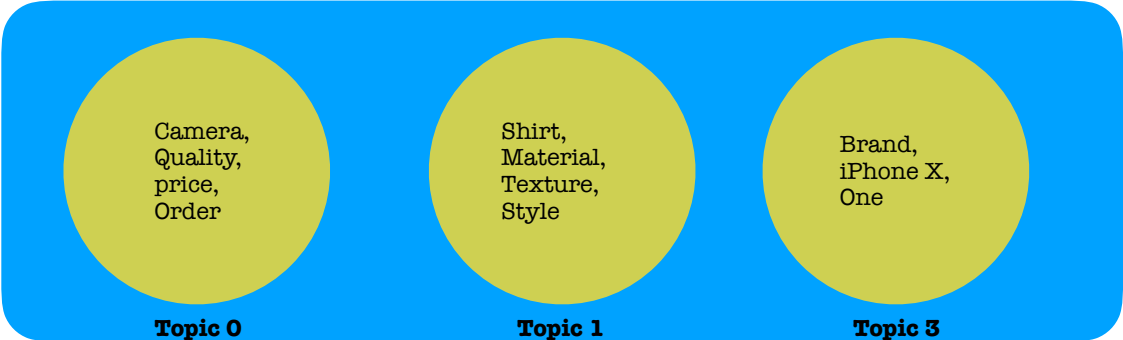
Review 4 : This shirt has very good material. Texture and style is what makes this brand awesome.

Review 1 : Iphone x, camera, one, store, home, features.

Review 2 : Samsung s9, phone, sound quality, build

Review 3 : Hello guys, shirt, Amazon, price, kidney, order

Review 4 : shirt, material, Texture, style, brand

Camera, Quality, price, Order

Shirt, Material, Texture, Style

Brand, iPhone X, One

Topic 0   **Topic 1**   **Topic 3**

1.Data Collection

	product_description	meta_description	description	bread_crumb	highlights	formated_reviews
0	NaN	Buy Moto G5s (Lunar Gray, 32 GB) online at bes...	NaN	Home > Mobiles & Accessories > Mobiles > Motor...		['Nice Good but over priced', 'Absolute rubbis...
1	NaN	Buy Imported ShengShow Gold Mirror Rubik Magic...	Very Good Fantastic Product. No compromise on ...	Home > Toys > Puzzles & Board Games > Puzzles ...		['Really Nice Smooth', 'Excellent It is a good...
2	A revolutionary camera that adapts like the hu...	Buy Samsung Galaxy S9 (Midnight Black, 128 GB)...	Galaxy S9 Sleek and stunning the latest flagsh...	Home > Mobiles & Accessories > Mobiles > Samsu...		['Must buy! Just woww', 'Fakruddin Nice Produc...

```
(266242, 6)
```

2. Feature engineering

1. **Outlier**
   Problems that we might face with such datasets are:
   1. How to find outliers?
   2. How to deal with these outliers?

2. **Missing data**
   Problems that we might face with such datasets are:
   1. How to find missing data?
   2. How to deal with these missing data?

3. **Different types of variables**
   Problems that we might face with such datasets are:
   1. What are the different types of variables/data?
   2. How to deal with different types of variables?

4. **Irrelevant data**
   Problems that we might face with such datasets are:
   1. How to identify irrelevant data?
   2. How to deal with irrelevant data?

5 **Extract Nouns**

6 **TFIDF**

7 **NMF**

Perform feature engineering on all 6 columns

product_description	meta_description	description	bread_crumb	highlights	formated_reviews
0 NaN	Buy Moto G5s (Lunar Gray, 32 GB) online at bes...	NaN	Home > Mobiles & Accessories > Mobiles > Motor...	▯	['Nice Good but over priced', 'Absolute rubbis...
1 NaN	Buy Imported ShengShow Gold Mirror Rubik Magic...	Very Good Fantastic Product. No compromise on ...	Home > Toys > Puzzles & Board Games > Puzzles ...	▯	['Really Nice Smooth', 'Excellent It is a good...
2 A revolutionary camera that adapts like the hu...	Buy Samsung Galaxy S9 (Midnight Black, 128 GB)...	Galaxy S9 Sleek and stunning the latest flagsh...	Home > Mobiles & Accessories > Mobiles > Samsu...	▯	['Must buy! Just woww', 'Fakruddin Nice Produc...

Feature engineering in text

Nouns   **Nouns**   **Nouns**   **Nouns**   **Nouns**   **Nouns**

Review 1 : Iphone x, camera, one, store, home, features.

Review 2 : Samsung s9, phone, sound quality, build

Review 3 : Hello guys, shirt, Amazon, price, kidney, order

Review 4 : shirt, material, Texture, style, brand

5 TF IDF

6 NMF

Attribute detection solution

Thank you