

The background is white and decorated with various watercolor-style elements. There are several large, overlapping circles in shades of teal, yellow, and orange-red. Scattered throughout are smaller yellow stars, some with radiating lines, and small orange starburst shapes. The overall aesthetic is bright and cheerful.

Classifying happiness

Summer Yuan 2017/2/14



1. Data

2. EDA

3. Model

4. Test case

**5. Performance
& Findings**

标题文本预设

The background is a light cream color, decorated with various watercolor-style elements. There are several large, overlapping circles in shades of yellow, light blue, and pink. Scattered throughout are small, stylized starburst shapes in yellow and orange. The overall aesthetic is soft and celebratory.

1. Data

**Over 13000 individuals
from more than 84 countries
6 happy moments**

joined with world happiness report data

The background is white with various decorative elements. There are several large, semi-transparent circles in shades of yellow, light blue, and light pink. Scattered throughout are small, stylized starburst shapes in yellow and orange. Some circles have a darker, textured border or a different color at the bottom, like a green or orange band.

2. EDA

1. Does gender/country origin affect happiness category?

2. What contributes to happiness on country level?

3. Model - overview

Y ~ emotion category	
affection	achievement
bonding	enjoy-the-moment
leisure	nature
exercise	

X		
happy score	gdp	age
family	health	gender
freedom	generosity	marital
exercise	corruption	parenthood

3. Model -base model

	accuracy	f1	precision	recall
RandomForest	0.400617	0.358700	0.384483	0.400617
GaussianNB	0.267395	0.275642	0.309759	0.267395
SVC	0.199240	0.236378	0.366468	0.199240
Kneighbors	0.362384	0.321559	0.311473	0.362384
LogisticRegression	0.386844	0.305956	0.253587	0.386844

3. Model -improvement

1

**use
vectorized
text
as feature**

2

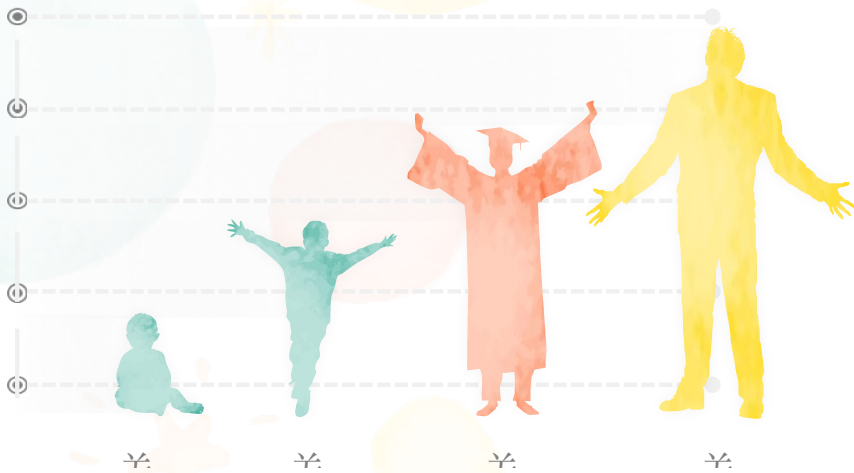
**smote
sample**

3

grid search

3. Model -adding vectorized text

```
from sklearn.feature_extraction.text import CountVectorizer  
from sklearn.decomposition import NMF
```



Topic #0: and, together, the, to, spend, night, other, each, families, play

Topic #1: my, house, mom, favorite, dad, dog, girlfriend, boss, office, sister

Topic #2: the, were, by, same, past, park, took, way, weather, sun

3. Model -adding vectorized text

Random forest accuracy score

0.4 >> 0.76

	accuracy	f1	precision	recall
RandomForest	0.765851	0.750535	0.765709	0.765851
GaussianNB	0.367371	0.413211	0.647437	0.367371
SVC	0.184042	0.202590	0.353696	0.184042
Kneighbors	0.460936	0.439562	0.446052	0.460936
LogisticRegression	0.586084	0.526551	0.614491	0.586084

3. Model -smote sampling

Random forest accuracy score

0.4 >> 0.76

	precision	recall	f1-score	support
achievement	0.69	0.85	0.76	1282
affection	0.83	0.89	0.86	1466
bonding	0.90	0.81	0.85	514
enjoy_the_moment	0.62	0.40	0.49	431
exercise	0.87	0.20	0.33	64
leisure	0.76	0.58	0.66	379
nature	0.69	0.12	0.20	75
avg / total	0.77	0.77	0.75	4211

3. Model -smote sampling

Random forest accuracy score

0.4 >> 0.76>>0.73

	precision	recall	f1-score	support
achievement	0.69	0.77	0.73	1282
affection	0.81	0.83	0.82	1466
bonding	0.85	0.81	0.83	514
enjoy_the_moment	0.45	0.42	0.43	431
exercise	0.52	0.25	0.34	64
leisure	0.70	0.55	0.62	379
nature	0.41	0.31	0.35	75
avg / total	0.72	0.73	0.72	4211

3. Model -randomized search

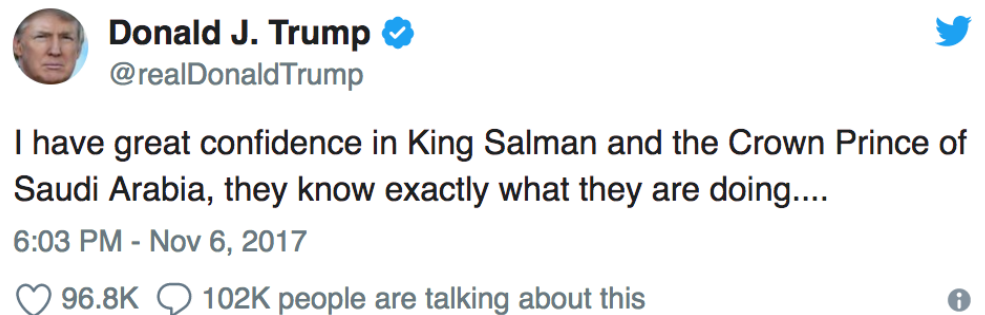
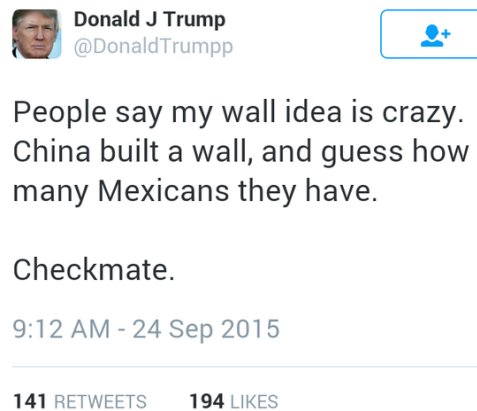
Random forest accuracy score

0.4 >> 0.76>>0.73>>0.81

```
{'max_depth': 30,  
'max_features': 'sqrt',  
'min_samples_leaf': 1,  
'min_samples_split': 2,  
'n_estimators': 2000}
```

	precision	recall	f1-score	support
achievement	0.77	0.87	0.81	1282
affection	0.86	0.92	0.89	1466
bonding	0.93	0.84	0.88	514
enjoy_the_moment	0.64	0.50	0.56	431
exercise	0.82	0.22	0.35	64
leisure	0.80	0.70	0.75	379
nature	0.62	0.31	0.41	75
avg / total	0.81	0.81	0.80	4211

4. Test case : try it out!



4. Test case :

Result: {'achievement': 100}

Top topics that makes trump happy...

#['all', 'make', 'again', 'america', 'great', 'support', 'rallies', 'thanks', 'today', 'trump']

make america great again, thanks to Trump

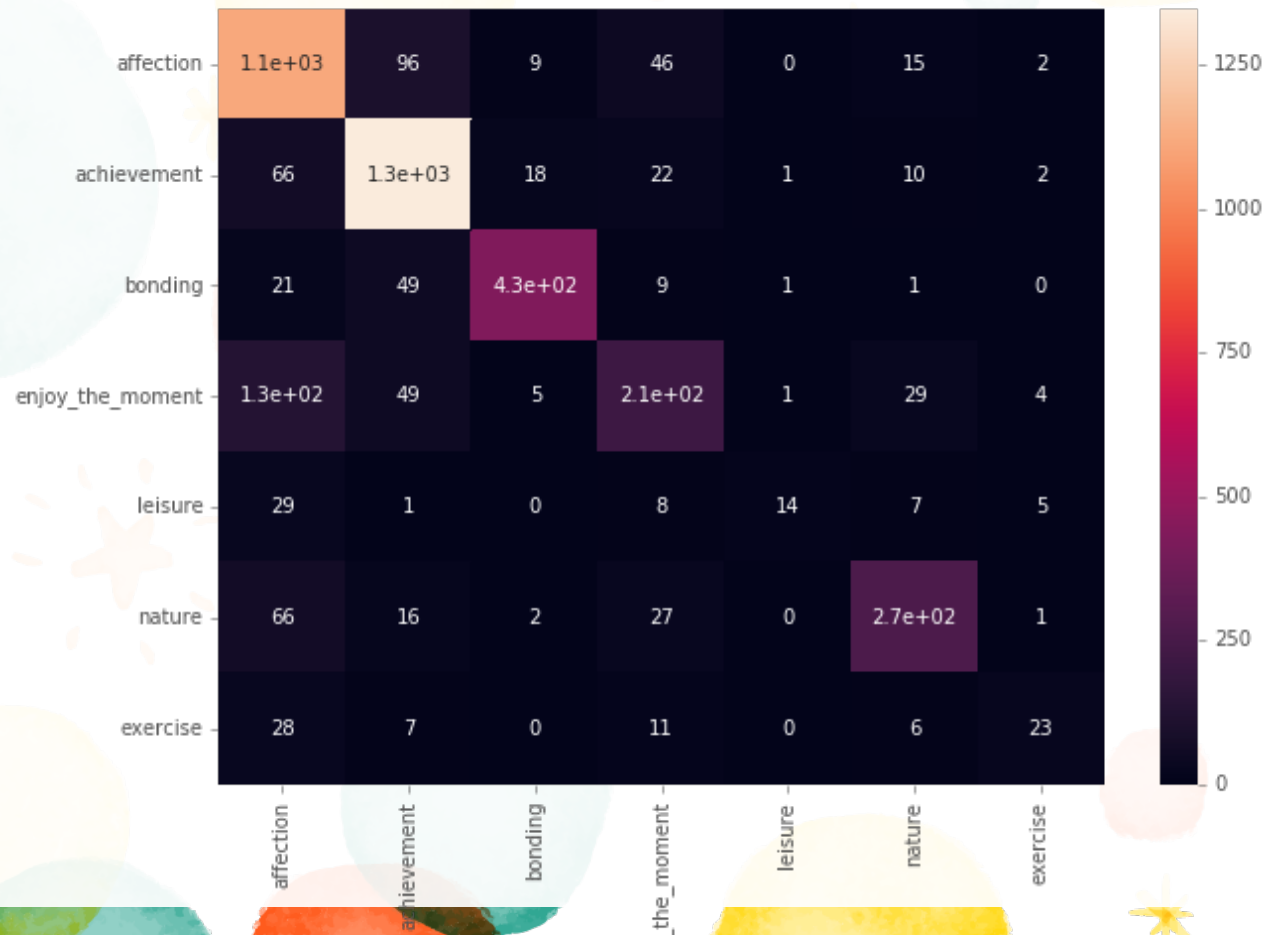
#['the', 'negative', 'news', 'it', 'but', 'is', 'economy', 'record', 'mainstream', 'Russia']

despite negative news on Russia, the economy is on record

#['zoo', 'greet', 'governor', 'good', 'golf', 'going', 'goes', 'goal', 'go', 'gives']

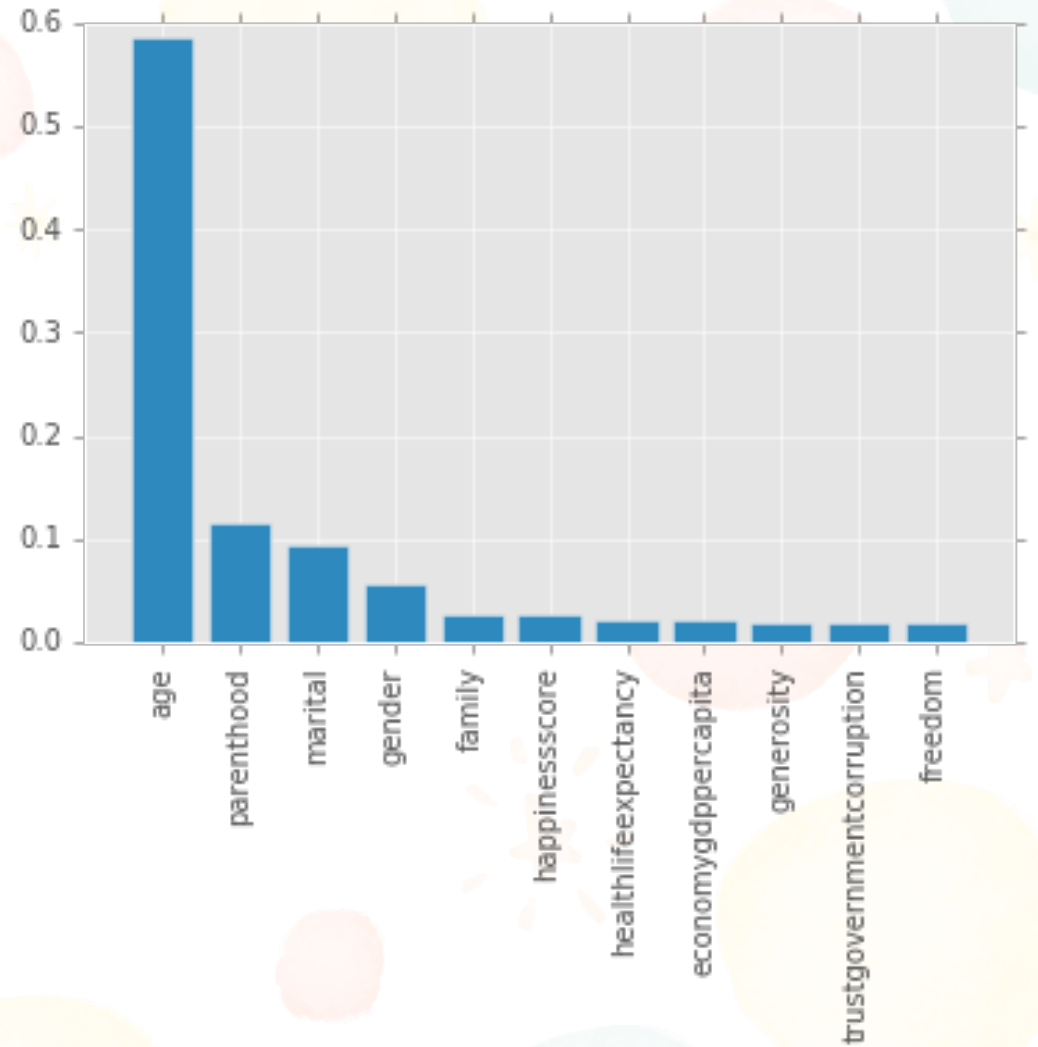
greet the governor, go golfing, go to the zoo???

5. Evaluation & Result



5.Evaluation & Result

feature importance before adding
vectorized features



5.Evaluation & Result

1 ['not', 'wedding', 'they', 'aunts', 'uncles', 'but', 'his', 'if', 'bride', 'married']

2 ['to', 'and', 'the', 'do', 'us', 'who', 'help', 'also', 'people', 'needed']

3 ['all', 'are', 'members', 'there', 'together', 'always', 'll', 'ceremony', 'passed', 'organized']

4 ['see', 'them', 'again', 'who', 'seen', 'dog', 'little', 'coming', 'haven', 'visit']

5 ['wanted', 'just', 'her', 'life', 'grandma', 'as', 'really', 'did', 'has', 'years']

>>

1 celebrate life milestone with family

2 help those in need

3 spend time with your loved ones

4 regain what was lost

5 fulfill family's last wishes



Thank you