

Aspiring Minds Research Team (research.aspiringminds.com)

Twitter sentiment analysis

What is Python?

- General purpose interpreted programming language
- Widely used by scientists and programmers of all stripes
- Supported by many 3rd-party libraries (currently 21,054 on the main python package website)
- **Libraries** in python are known as **packages**.
- You can install particular library using **pip** – this is popular default package manager.
 - Eg – pip install numpy

Why is it well-suited to science?

- **Some useful packages for data science**
- Numpy
 - Numerical library for python
 - Written in C, wrapped by python
- Scipy
 - Built on top of numpy (i.e. Also fast!)
 - Common maths, science, engineering routines
- Matplotlib
 - Hugely flexible plotting library
- Pandas
 - Easy to use library for data analysis and modelling
 - Dataframes are like sql tables and you can perform operations
 - Read, write, data reshaping, slicing, joins, pivots, indexing, etc

iPython notebook – quick intro

- IPython Notebook is a web-based interactive computational environment
- There is a notebook which runs in browser and it connects to a server having a kernel
- Containing an ordered list of input/output cells which can contain code, text, mathematics, plots and rich media.
- How to start ipython notebook
 - `$ ipython notebook`

Hands on – big picture

- Steps

- 1. Setup your ipython environment and download libraries
- 2. Setup API keys for your twitter account
- 3. Fetch tweets from twitter
- 4. Run sentiment analysis
- 5. Analyze results

Set up ipython

- Clone the repo
 - git clone <https://github.com/gursimar/iPython-workbooks.git>
- Start your ipython notebook
 - Using ipython notebook
 - Goto - www.tmpnb.org
- Upload the **ipynb file** from the repo
 - twitter_sentiment_analysis.ipynb
- Install if any packages are required using **pip**

Play with ipython (CP₁)

- It contains cells which may have text/ code
- You can add a cell by using the **+** **icon** above
- Remove a cell using the **scissor icon**.
- **Shift+Enter** to execute code inside the cell
- Exercise → add a new cell and print 'hello'
- Exercise 2 -> + cell and

Setup twitter api keys (CP2)

- Open <https://apps.twitter.com/>
- Click "Create New App"
- Fill out the form, agree to the terms, and click "Create your Twitter application"
- Click on "Keys and Access Tokens" tab, and copy your "API key" and "API secret".
- Scroll down and click "Create my access token", and copy your "Access token" and "Access token secret".
- Access Level -> Read, write, and direct messages (modify app permissions)
- - <https://apps.twitter.com/app/13519590/keys>

Twitter API

- OAuth is used to authenticate and authorize
- Twitter API Objects
 - Tweets
 - Users
 - Entities – hashtags, media, urls, etc
 - Places

python-twitter - wrapper

- It is a python wrapper to the twitter API
- Create an object
 - `api = twitter.Api(credentials)`
- Interact with twitter API through functions
 - `Api.GetFriends()`
 - `Api.PostUpdate()`
 - `Api.GetSearch()`
 - `Api.GetTrendsCurrent()`
 - etc

Sentiment analysis overview

- It's the voice of the customer
 - What are people feeling about a particular topic/product
- Classify the polarity of the article
 - Positive, negative, neutral
- Various techniques

Sentiment analysis – a basic approach

- Example
- Tokenize and remove stop words
- Remove Stop words
- POS Tagging – select only adjectives/ adverbs
- Pass these tokens to a pretrained sentiment classifier

Sentiment analysis – classifier training

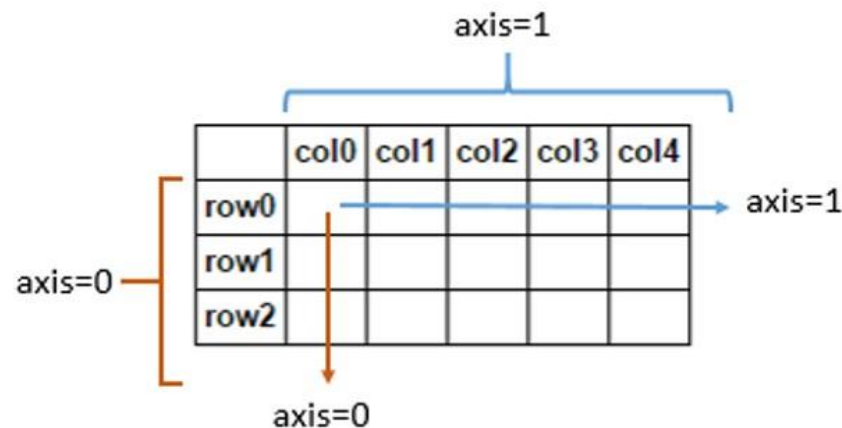
- Trained using a supervised data set
 - Movie review dataset
 - Each review has a label – positive or negative
- We process the text as in the previous slide
- Then we use a bag of words approach
- These features, along with labels are passed to a Naïve bayes classifier.
- It gives a score on a scale of $[-1, 1]$

TextBlob python package

- It is a package to process textual data.
 - NLP, POS tagging, Noun phrase extraction, sentiment analysis, classification, translation.
- `analysis = TextBlob(clean_tweet(tweet))`
- `if analysis.sentiment.polarity > 0:`
 - `return 'positive'`
- `elif analysis.sentiment.polarity == 0:`
 - `return 'neutral'`
- `else:`
 - `return 'negative'`

Pandas data frame

- It's a python package for data manipulation and analysis
 - Mainly useful to work on tabular data like in spreadsheets.
- A data frame is used for storing data tables



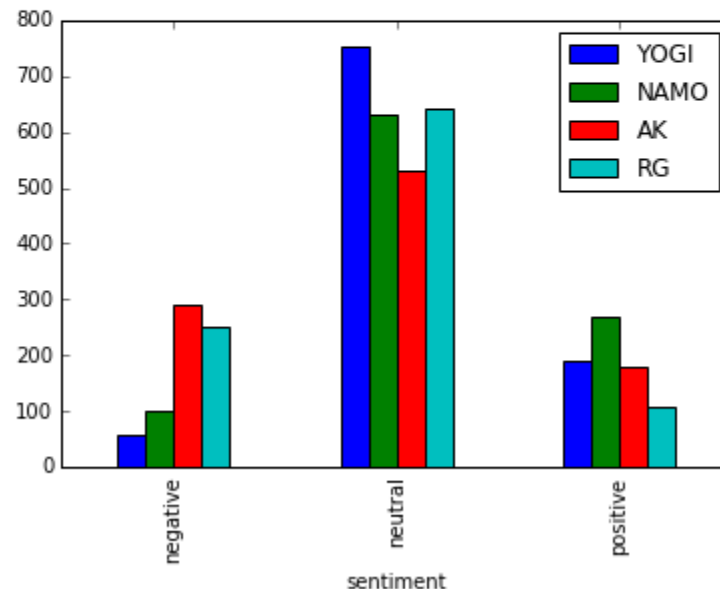
Pandas data frame

- One line commands can be performed
- Features – data manipulation
 - Loading/ dumping data from/to csv
 - `pd.from_csv('simar.csv')`
 - Joins and data merging/ alignment
 - Reshaping and pivoting data sets
 - Handling missing data
 - Advanced indexing and slicing of data

Analyze results and draw graphs

	YOGI	NAMO	AK	RG
sentiment				
negative	58	99	290	251
neutral	753	631	530	643
positive	189	270	180	106

Out[179]: <matplotlib.axes._subplots.AxesSubplot at 0x1f28eb70>



What next!!

■ Analyze data in excel

	B	C	D	E	F	G	H	I	J	K
	favorite_count	retweet_count	sentiment	source	text	user_fav_count	user_follower	user_friends_count	user_name	user_screen_name
1	0	21758	neutral	<a href="http://t.co/...	RT @Gurme...	2097	4	8	Ajita Bansal	ajitainsaan
4	0	21758	neutral	<a href="http://t.co/...	RT @Gurme...	2097	4	8	Ajita Bansal	ajitainsaan
9	0	21758	neutral	<a href="http://t.co/...	RT @Gurme...	2097	4	8	Ajita Bansal	ajitainsaan
5	0	21758	neutral	<a href="http://t.co/...	RT @Gurme...	2097	4	8	Ajita Bansal	ajitainsaan
7	0	21760	neutral	<a href="http://t.co/...	RT @Gurme...	2097	4	8	Ajita Bansal	ajitainsaan
0	0	21761	neutral	<a href="http://t.co/...	RT @Gurme...	2097	4	8	Ajita Bansal	ajitainsaan
3	0	21761	neutral	<a href="http://t.co/...	RT @Gurme...	2097	4	8	Ajita Bansal	ajitainsaan
7	0	21763	neutral	<a href="http://t.co/...	RT @Gurme...	2097	4	8	Ajita Bansal	ajitainsaan
1	0	21764	neutral	<a href="http://t.co/...	RT @Gurme...	2097	4	8	Ajita Bansal	ajitainsaan
7	0	21764	neutral	<a href="http://t.co/...	RT @Gurme...	2097	4	8	Ajita Bansal	ajitainsaan

- Same user is doing multiple retweets
- Consider users with fav_count > th | user_follow > th
- Take only one tweet per user; consider retweet count
- See user stats rather than tweet stats
- Implement our own algorithm for sentiment analysis
- Get facebook posts along with twitter

The End