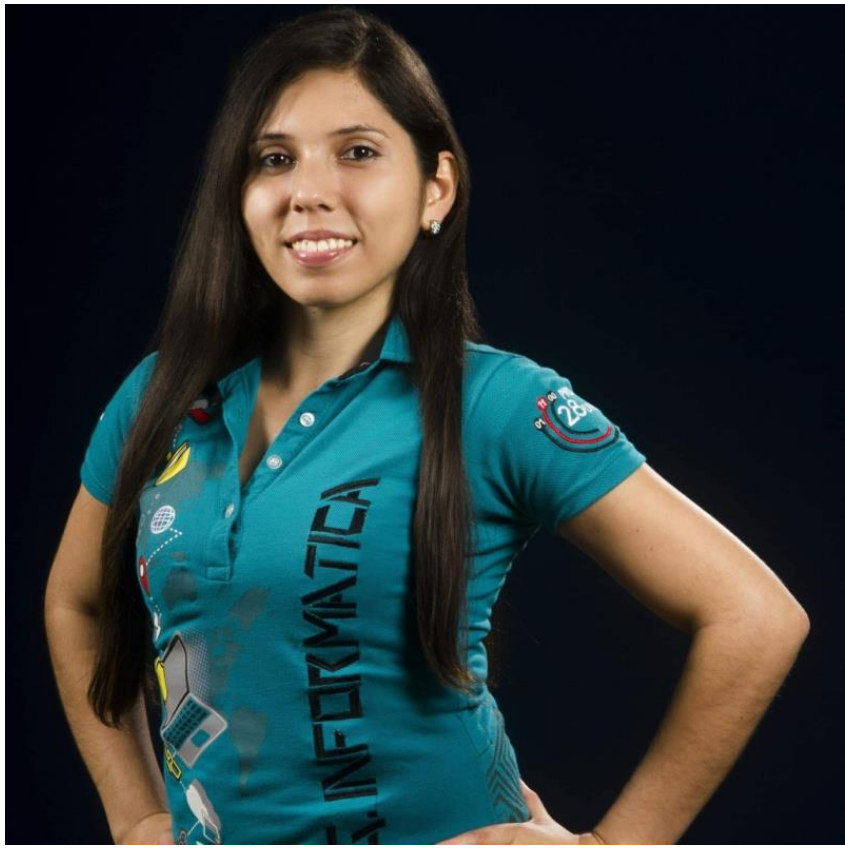
Hello Hello

#PythonPereira

Angelica Landazabal

- Computer Science Engineering and a technology lover. (Karateka too)
- I like the development of web applications, marketing and, of course, data analysis ;)
- It motivates me to continue learning about various IT topics, personal / professional growth and helping others so they are also motivated to learn.
- "Keep going" is my motto :)



angelica.informatik@gmail.com

ALandazabal

@ALandazabal15

django girls Colombia

PionerasDev

Hackem
Cybersecurity Research Group

FLISol

Platzi


SciPy
LATAM·2019
7ª CONFERENCIA
LATINOAMERICANA DE
PYTHON CIENTÍFICO
OCT 8 - 10, 2019 • BOGOTÁ, COLOMBIA • CONF.SCIPYLA.ORG

Lindsey Heagy

Travis Oliphant

Damián Avila

https://www.youtube.com/channel/UClb88lwUvlFikmhTzVGsVGA/featured

"Applying sentiment analysis to your social network"

What is Sentiment Analysis?

Objectives:

- Understand What is Natural Language Processing y which are its applications.
- Extract data from a Social Network.
- Cleaning the data extracted.
- Apply the Sentiment Analysis technique to 2 sets of data (Spanish - English).
- Show data processed.

What will we need?

https://github.com/ALandazabal/Python-Pereira---Analisis-de-Sentimiento

Natural Languaje Processing (NLP)

Is the study that it try to understand, analyze, interpret and obtain mean of human languaje with a computer to will be used in data science, AI and linguistics.

The main idea is perceive of a set of words the intention of its creator.

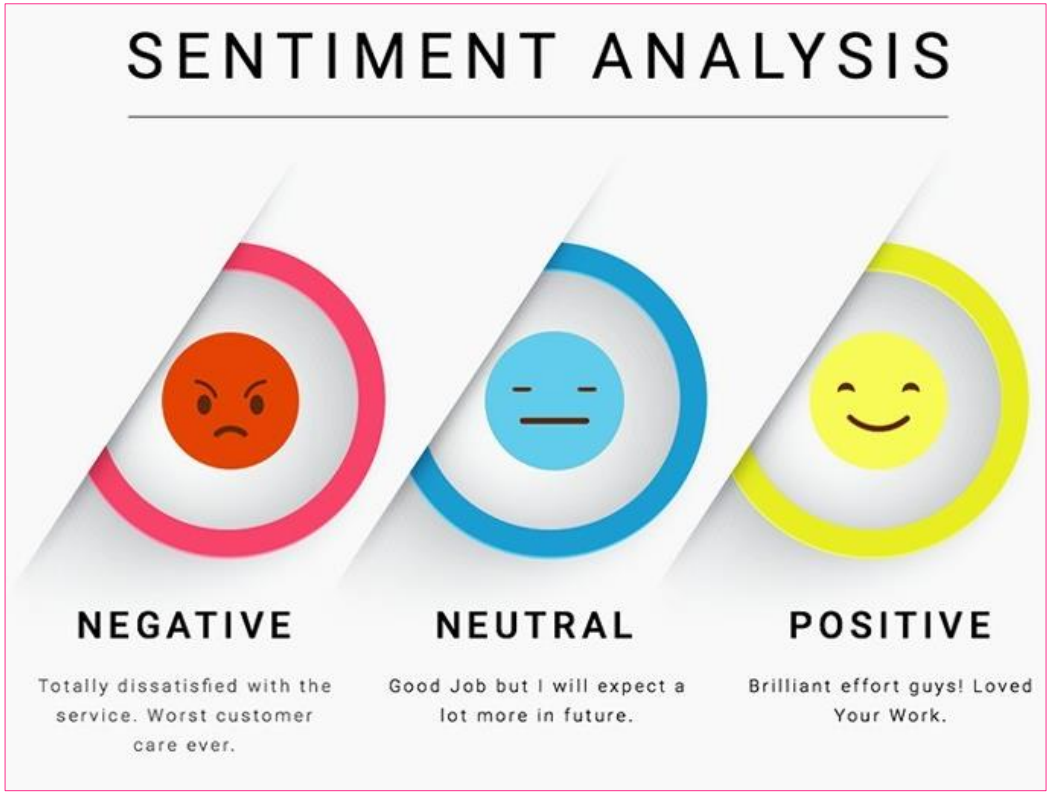
NLP applications:

- **Translation:** It is the most used application because it allows you to understand more quickly the text written in another language.

- **Automatic summarization:** the idea is get a reduced version of a text created by extraction or abstraction.

- **Chatbots:** A system capable of having a coherent conversation with a human about a specific topic.

- **Sentiment Analysis:** Identify subjective information in texts as judgments or opinions used for companies and famous people that want to be aware of their reputation on the internet.

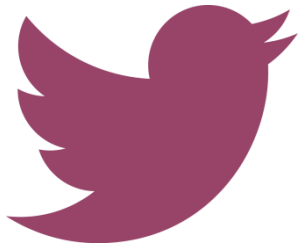


Steps for Sentiment Analysis:

Step 1: Extract Data

- **Social Network:** Twitter

Twitter is what is happening in the world and what people are talking about at the moment. You can access Twitter on the Web or from your mobile device.



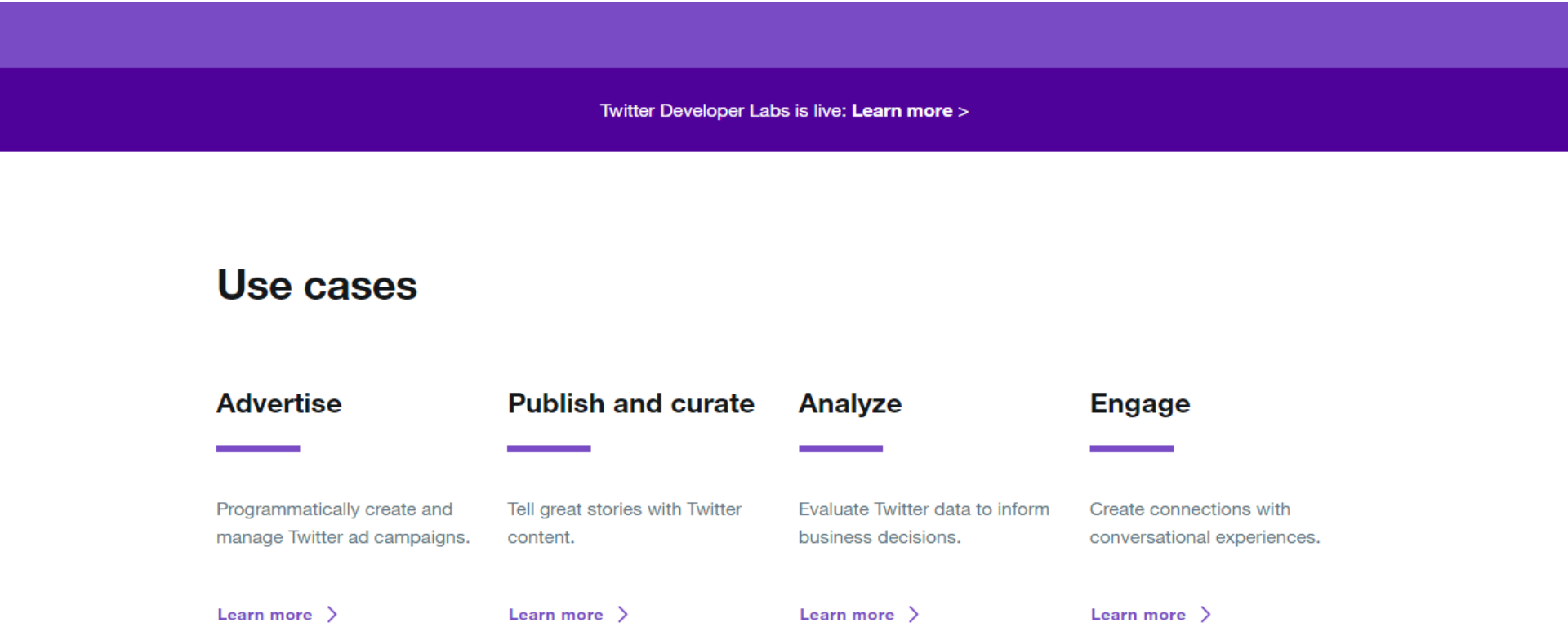
Step 1: Extract Data

- **API Twitter:** Twitter provides companies, developers and users programmatic access to data through its API (application programming interface).

  You should be a twitter developer account and create an app to request the access tokens.

  Note: We only access public information

Step 1: Extract Data



https://developer.twitter.com/en.html

Step 1: Extract Data

- **Tweepy:** An easy-to-use Python library for accessing the Twitter API.

  Install with :

  pip install tweepy

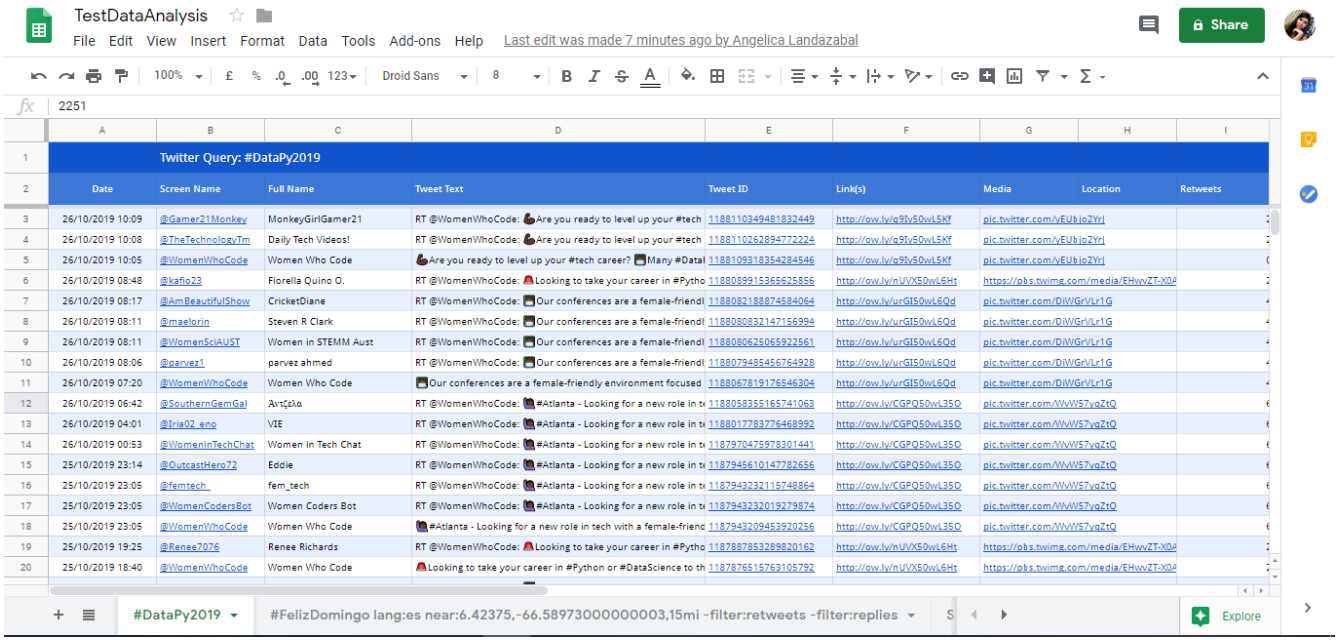
- **Csv:** Is a Python module for creating CSV files.

  Install with :

  pip install python-csv

Step 1: Extract Data

- **Twitter Archiver:** is an Add-ons in Drive Spreadsheet help us to extract data of twitter with the parameters given.



https://docs.google.com/spreadsheets/d/1tUlLulV5Y-yWivYHnLw4zWh7DEKDCC7DNVbtqPXEnZU/edit?usp=sharing

Step 1: Extract Data

- **Facebook-Scraper:** Scrape Facebook public pages without an API key. Inspired by twitter-scraper.

  Install with :

  pip install facebook-scraper

Step 2: Load Data

- **Pandas:** Powerful data structures for data analysis, time series, and statistics.

  Install with : pip install pandas

- **NLTK:** The Natural Language Toolkit (NLTK) is a Python package for natural language processing.

  Install with : pip install nltk

Step 3: Clean Data

- **String:** contains some constants, utility function, and classes for string manipulation.

  Install with : pip install strings

- **Re:** This module provides regular expression matching operations similar to those found in Perl.

  Install with : pip install re

- **Emoji:** The entire set of Emoji codes as defined by the unicode consortium is supported in addition to a bunch of aliases.

  Install with : pip install emoji

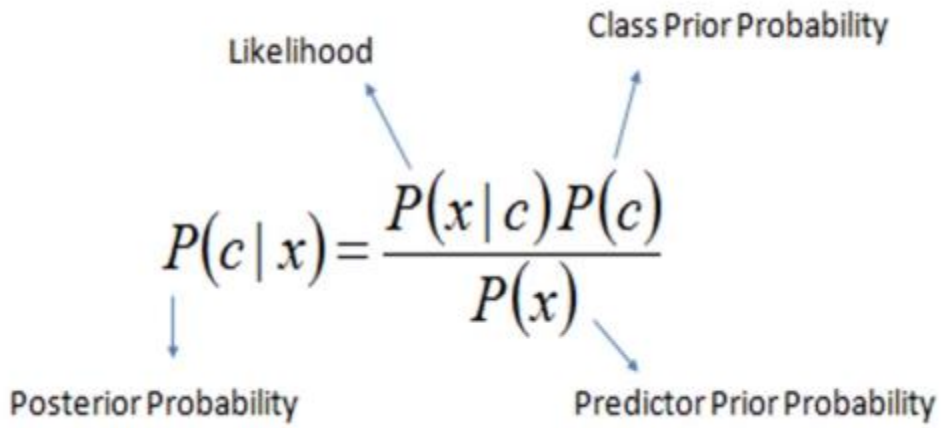
Objectives:

- Understand What is Natural Language Processing y which are its applications. ✓
- Extract data from a Social Network. ✓
- Cleaning the data extracted. ✓
- Apply the Sentiment Analysis technique to 2 sets of data (Spanish - English).
- Show data processed.

,

Step 4: Processing Data

- **Naive Bayes:** It is a classification and prediction technique that constructs models to determine what size is the possibility in which a set of data originates certain results..



,

Step 4: Processing Data

Total	Positive		Negative	
100	70		30	
Learn	55	0,78	5	0,16
Patterns	40	0,57	10	0,33
…	…	…	…	…
Learn and Patterns	31	0,44	1,5	0,05

$$P(Pos \mid \text{Learn and Patterns}) = \frac{31}{31 + 1{,}5} = 0{,}95 \Rightarrow 95\%$$

$$P(Neg \mid \text{Learn and Patterns}) = \frac{1{,}5}{31 + 1{,}5} = 0{,}05 \Rightarrow 5\%$$

,

Step 4: Processing Data

- **Bag of word:** It is a mathematical representation of the testing data set. This has all the unique words of a document and the occurs frequency of each one.

	it	is	puppy	cat	pen	a	this
it is a puppy	1	1	1	0	0	1	0
it is a kitten	1	1	0	0	0	1	0
it is a cat	1	1	0	1	0	1	0
that is a dog and this is a pen	0	2	0	0	1	2	1

Step 5: Show Data

- **Matplotlib:** is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.

  Install with: pip install matplotlib

- **WordCloud:** This tool will be quite handy for exploring text data and making your report more lively.

  Install with: pip install wordcloud

Processing Data (English text)

- **Textblob:** It provides a consistent API for diving into common NLP tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, and more.

Sentiment Analysis returns two values:

Polarity: float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement.

Subjectivity: Subjective sentences generally refer to personal opinion. Subjectivity is also a float which lies in the range of [0,1].

Install with: pip install textblob

Objectives:

- Understand What is Natural Language Processing y which are its applications. ✓
- Extract data from a Social Network. ✓
- Cleaning the data extracted. ✓
- Apply the Sentiment Analysis technique to 2 sets of data (Spanish - English). ✓
- Show data processed. ✓

Use cases of Sentimental Analysis

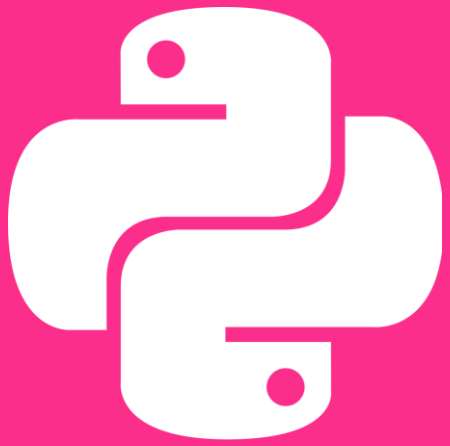
- Brand monitoring
- Competitive research
- Flame detection and customer service prioritization
- Product analysis
- Market research and insights into industry trends
- Workforce analytics/employee engagement monitoring

Sources

- **The Definitive Guide to Natural Language Processing :** https://monkeylearn.com/blog/definitive-guide-natural-language-processing/
- **API Twitter: https:** //developer.twitter.com/
- **Jupyter Notebook:** https://jupyter.org/
- **Excel files with python:** https://xlsxwriter.readthedocs.io/
- **Tweepy Library:** https://www.tweepy.org/
- **Pandas: https:** //pandas.pydata.org/
- **NLTK: https:** //www.nltk.org/
- **Re: https:** //docs.python.org/3/library/re.html

Sources

- **String module:** https://www.journaldev.com/23788/python-string-module
- **TextBlob: https:** //textblob.readthedocs.io/en/dev/
- **Matplotlib:** https://matplotlib.org/
- **WordCloud:** https://www.datacamp.com/community/tutorials/wordcloud-python
- **Sentiment Analysis: Types, Tools and Use Cases:** https://www.altexsoft.com/blog/business/sentiment-analysis-types-tools-and-use-cases/



“Las personas mienten, los datos no”
PlatziConf 2019

angelica.informatik@gmail.com

ALandazabal

@ALandazabal15