RESEARCH GROUP BISITE .usal.es

An AI Feature-Based Approach on Behavior Analysis of the users on Twitter

Niloufar Shoeibi

University of Salamanca

800 years of knowledge behind us

PyData SALAMANCA

VNiVERSiDAD D SALAMANCA CAMPUS OF INTERNATIONAL EXCELLENCE

800 AÑOS VNiVERSiDAD D SALAMANCA

IOT DIGITAL INNOVATION HUB

Introduction

Ph.D. in Intelligent Systems
@ the University of Salamanca(USAL), Spain

R&D in Artificial intelligence
@ BISITE Research Group

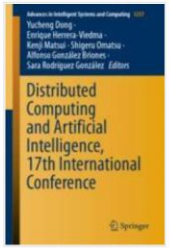
My research interests:

- social media analytics
- Data Science
- Artificial Intelligence
- Machine Learning
- Natural Language Processing
- Discreet Mathematics and Graph Analytics

And so on…

Niloufar Shoeibi
Niloufar.Shoeibi@USAL.es
@NilouShb

Springer Link



International Symposium on Distributed Computing and Artificial Intelligence
DCAI 2020: Distributed Computing and Artificial Intelligence, 17th International Conference pp 284-294 | Cite as

A Feature Based Approach on Behavior Analysis of the Users on Twitter: A Case Study of AusOpen Tennis Championship

Authors | Authors and affiliations

Niloufar Shoeibi, Alberto Martín Mateos, Alberto Rivas Camacho, Juan M. Corchado

Conference paper
First Online: 07 August 2020

Part of the Advances in Intelligent Systems and Computing book series (AISC, volume 1237)

Abstract

Due to the advancement of technology, and the promotion of smartphones, using social media got more and more popular. Nowadays, it has become an undeniable part of people's lives. So, they will create a flow of information by the content they share every single moment. Analyzing

OutLine:

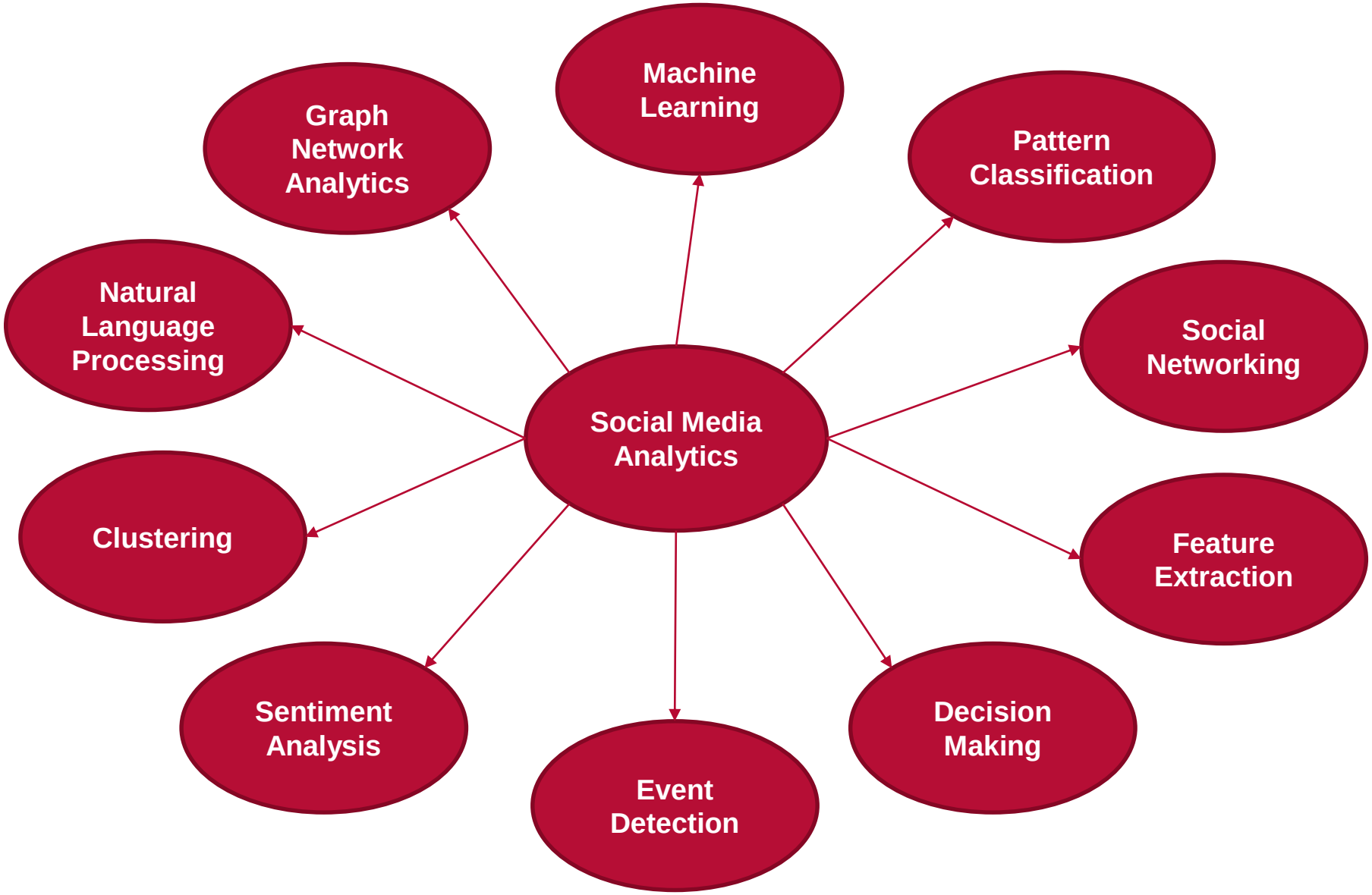
- What's Social Media Analytics?
- Research Areas & Practical Required Information
- Proposed Architecture & Case Study of US Election 2020
- Questions & Answers

What's Social Media Analytics?

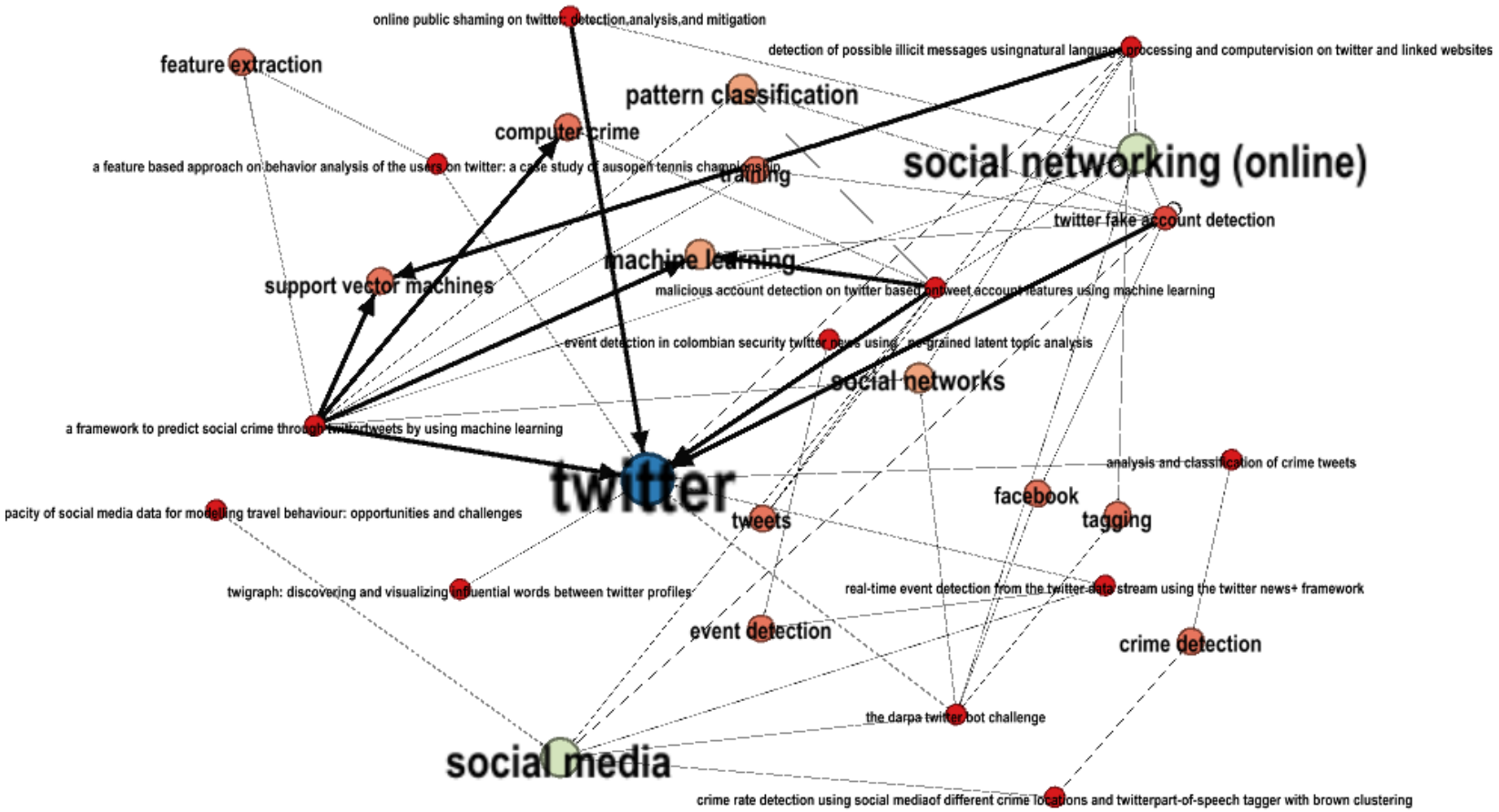
- Social Media Analytics is the process of gathering and analyzing data from social media platforms such as Facebook, Instagram, LinkedIn, Twitter and so on.

- The art and science of extracting hidden concepts from the Semi-Structured data, to enable informed and insight decision making.

Research Areas



Research Areas



Practical Required Information

- Privacy & Policy
- Data Publication Poclicies
- Building Twitter APP
- Tweet Object
- Data Extraction Methods

Privacy & Policy

- **User's Poclicy** → Location
- **Forbiden content**
- **The Algorithms Twitter contains** → Define your Project

Privacy & Policy

The Twitter Rules

help.twitter.com/en/rules-and-policies/twitter-rules

Help Center | Help topics | Guides | Contact us

Using Twitter
Managing your account
Safety and security
Rules and policies

- Twitter Rules and policies
- General guidelines and policies
- Law enforcement guidelines
- Research and experiments

↑ Scroll to top

Help Center > Twitter Rules and policies > The Twitter Rules

The Twitter Rules

Twitter's purpose is to serve the public conversation. Violence, harassment and other similar types of behavior discourage people from expressing themselves, and ultimately diminish the value of global public conversation. Our rules are to ensure all people can participate in the public conversation freely and safely.

Safety

Violence: You may not threaten violence against an individual or a group of people. We also prohibit the glorification of violence. Learn more about our violent threat and glorification of violence policies.

Terrorism/violent extremism: You may not threaten or promote terrorism or violent extremism. Learn more.

Child sexual exploitation: We have zero tolerance for child sexual exploitation on Twitter. Learn more.

Abuse/harassment: You may not engage in the targeted harassment of someone, or incite other people to do so. This includes wishing or hoping that someone experiences physical harm. Learn more.

Hateful conduct: You may not promote violence against, threaten, or harass other people on the basis of race, ethnicity, national origin,

Feedback

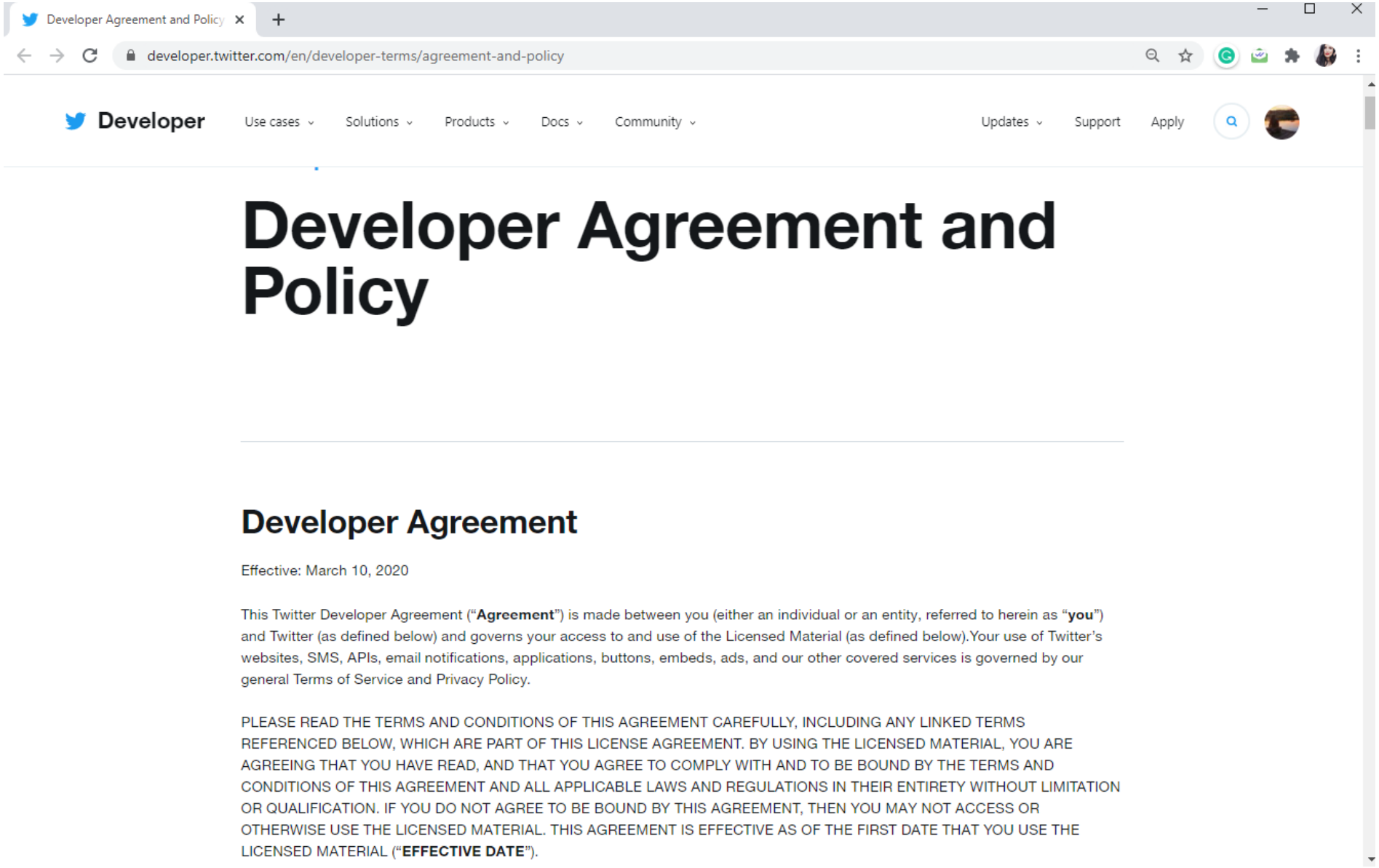
Privacy & Policy

Data Publication Poclicies

Data Publication Policies

Privacy & Policy

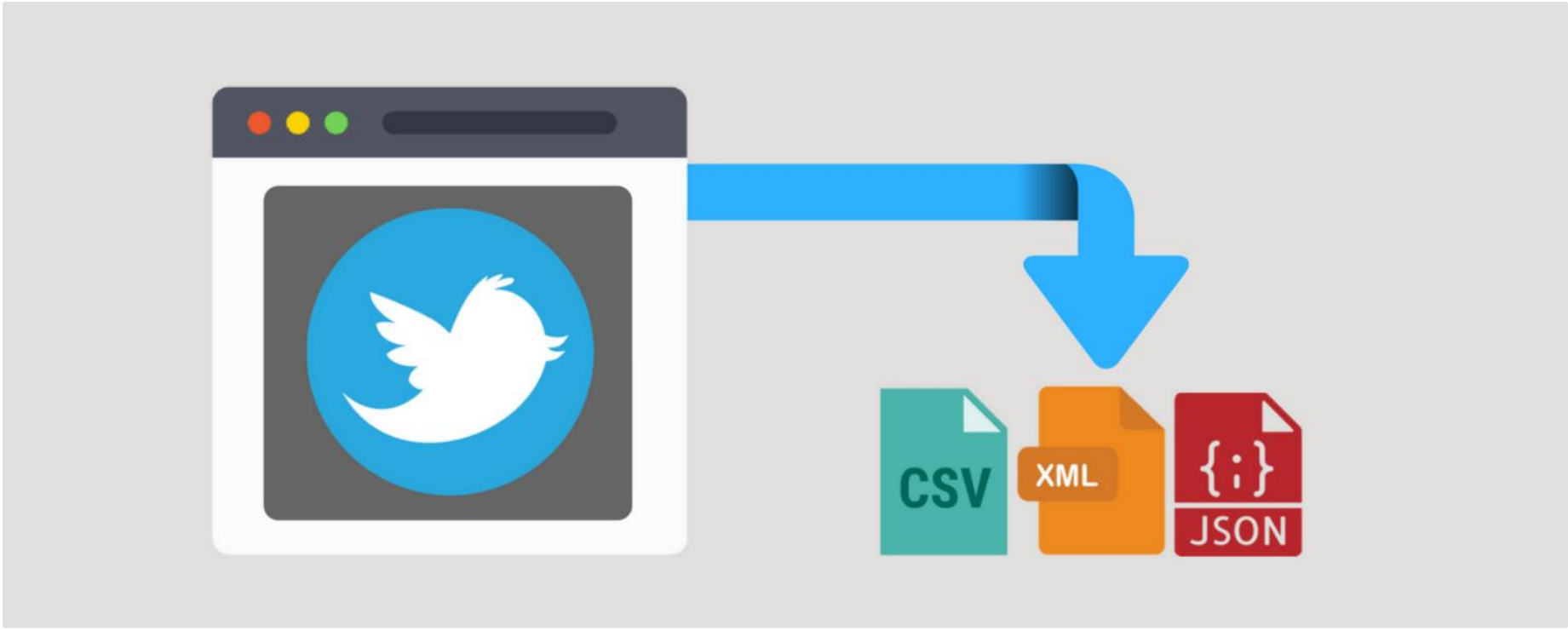
Data Publication Poclicies



Developer Agreement and Policy

- Privacy & Policy
- Data Publication Poclicies
- Building Twitter APP

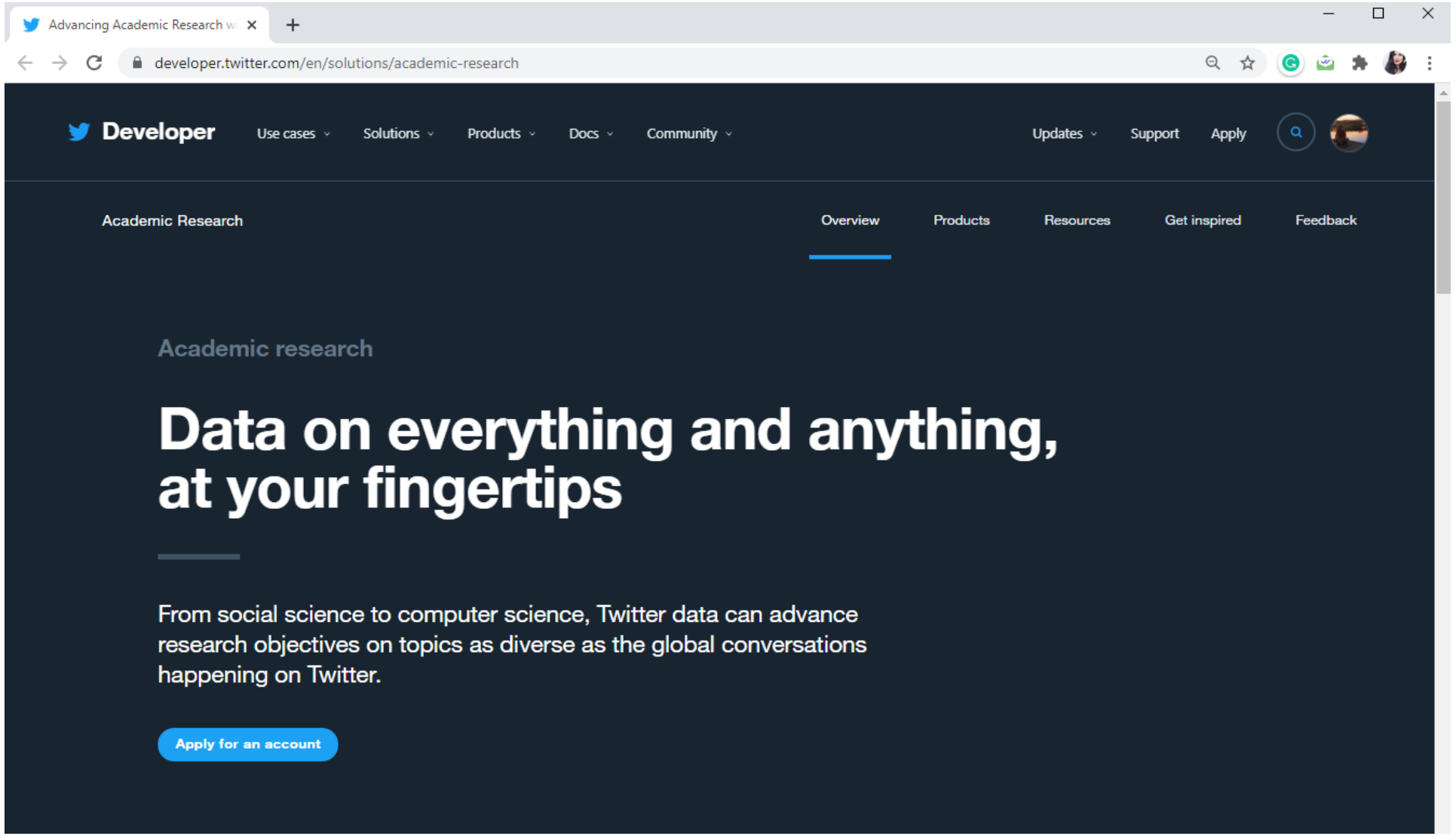
BuildingTwitter APP



Privacy & Policy

Data Publication Poclicies

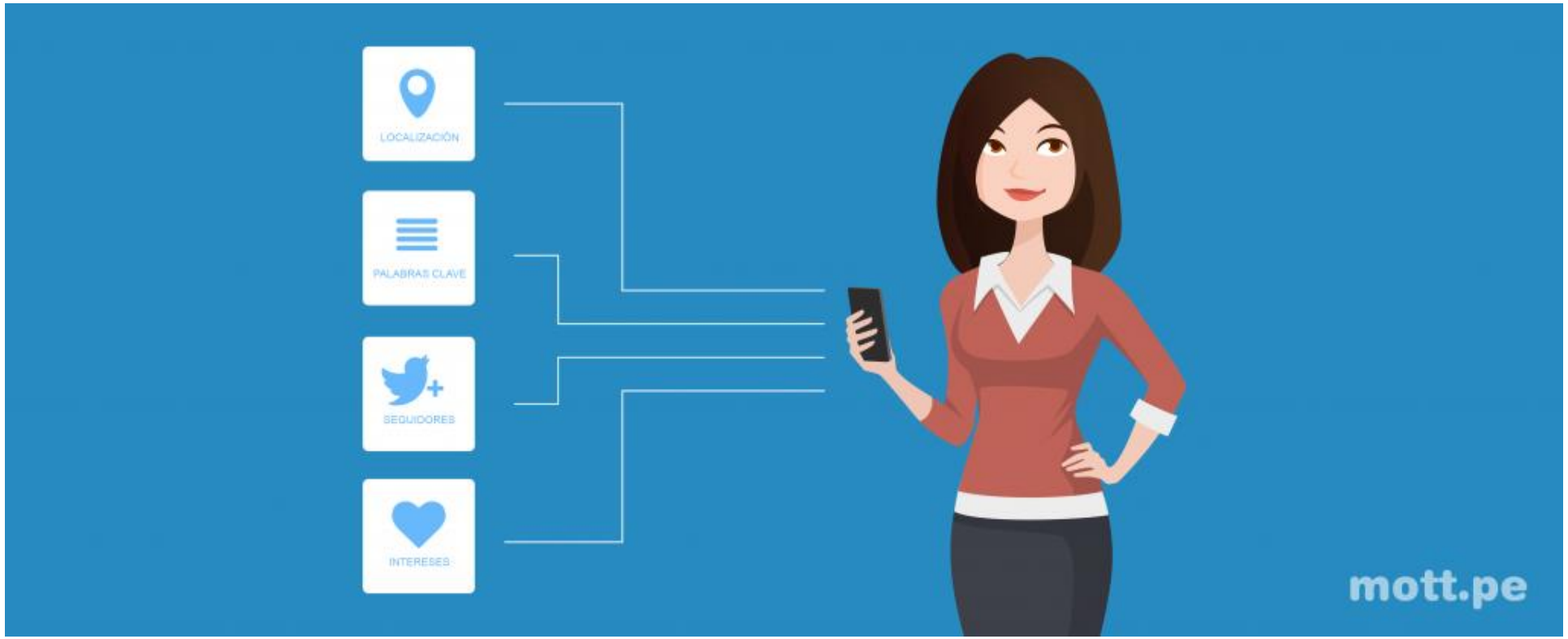
Building Twitter APP



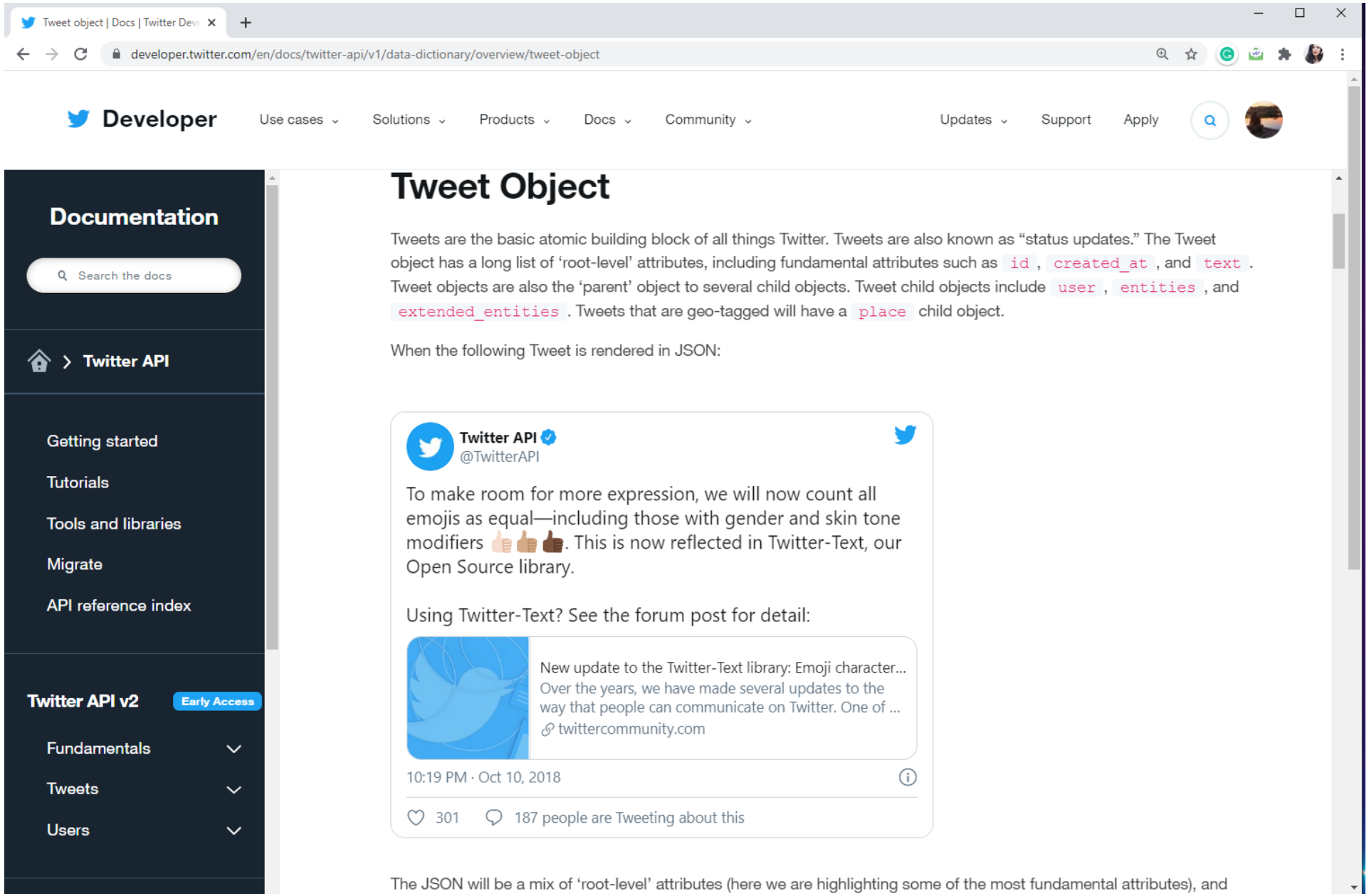
How to Build Twitter Credentials

- Privacy & Policy
- Data Publication Poclicies
- Building Twitter APP
- Tweet Object

Whats a Tweet Object?



- Privacy & Policy
- Data Publication Poclicies
- Building Twitter APP
- Tweet Object



Data Extraction Methods

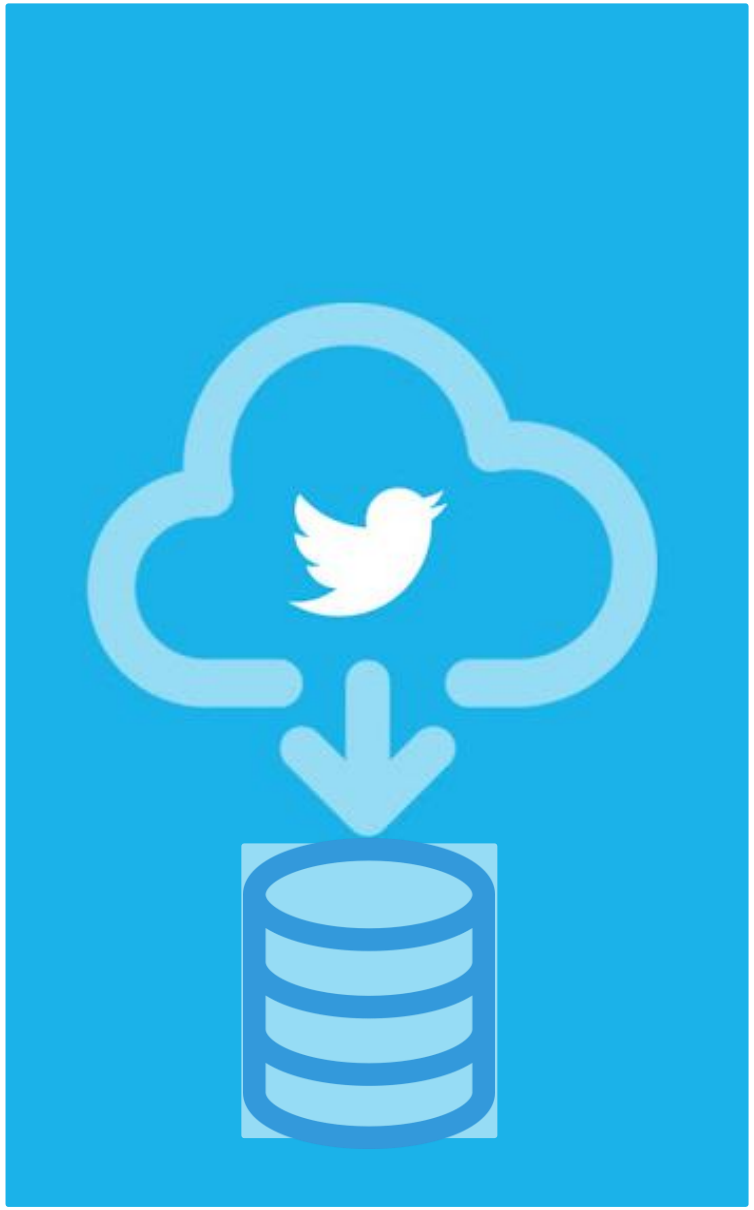
- Privacy & Policy
- Data Publication Poclicies
- Building Twitter APP
- Tweet Object
- Data Extraction Methods

Twitter APIs

- Twitter API v2: Early Access
- Twitter API v1.1
  - Standard APIs
  - Premium v1.1
  - Enterprise

Data Extraction Methods:

- **Real-Time:** Streaming the information
- **Batching Mode:** Downloading a specific amount of data all in one time.



- Privacy & Policy
- Data Publication Poclicies
- Building Twitter APP
- Tweet Object
- Data Extraction Methods

High-quality filters for getting Twitter data:

(Standalone Operators)

Operator	Description
Keyword	Matches a keyword within the body of a Tweet. This is a tokenized match, meaning that your keyword string will be matched against the tokenized text of the Tweet body.
Emoji	Matches an emoji within the body of a Tweet. Similar to a keyword, emojis are a tokenized match, meaning that your emoji will be matched against the tokenized text of the Tweet body.
”Exact Phrase Match”	Matches the exact phrase within the body of a Tweet.
#	Matches any Tweet containing a recognized hashtag, if the hashtag is a recognized entity in a Tweet.
@	Matches any Tweet that mentions the given username, if the username is a recognized entity (including the @ character).
From:	Matches any Tweet from a specific user.
To:	Matches any Tweet that is in reply to a particular user.
URL:	Performs a tokenized match on any validly-formatted URL of a Tweet.
retweets_of:	Matches Tweets that are Retweets of the specified user. The value can be either the username (excluding the @ character) or the user’s numeric user ID.

- Privacy & Policy
- Data Publication Poclicies
- Building Twitter APP
- Tweet Object
- Data Extraction Methods

High-quality filters for getting Twitter data:

(Standalone Operators)

Operator	Description
context:	NEW Matches Tweets with a specific domain id and/or domain id, enitity id pair. To learn more about this operator, please visit our page on annotations. context: domain_id.entity_id
entity:	NEW Matches Tweets with a specific entity string value. To learn more about this operator, please visit our page on annotations. entity:"string declaration of entity/place"
conversation_id:	NEW Matches Tweets that share a common conversation ID. A conversation ID is set to the Tweet ID of a Tweet that started a conversation. As Replies to a Tweet are posted, even Replies to Replies, the conversation_id is added to its JSON payload.

- Privacy & Policy
- Data Publication Poclicies
- Building Twitter APP
- Tweet Object
- Data Extraction Methods

High-quality filters for getting Twitter data:

(Non-Standalone Operators)

Operator	Description
is:retweet	Matches on Retweets that match the rest of the specified rule. This operator looks only for true Retweets (for example, those generated using the Retweet button). Quote Tweets will not be matched by this operator.
is:quote	Returns all Quote Tweets, also known as Tweets with comments.
is:verified	Deliver only Tweets whose authors are verified by Twitter.
has:hashtag	Matches Tweets that contain at least one hashtag.
has:links	This operator matches Tweets which contain links in the Tweet body.
has:mentions	Matches Tweets that mention another Twitter user.
has:media	Matches Tweets that contain a media URL recognized by Twitter.
has:images	Matches Tweets that contain a recognized URL to an image.
has:videos	Matches Tweets that contain native Twitter videos, uploaded directly to Twitter. This will not match on videos created with Periscope, or Tweets with links to other video hosting sites.
lang:	Matches Tweets that have been classified by Twitter as being of a particular language (if, and only if, the tweet has been classified).

An AI Feature-Based Approach on Behavior Analysis of the users on Twitter

Challenges:

- Stochastic and Dynamic Environment
- Non-stability of the features leading to not being able to have a proper dataset

  → Mapping the Problem to Graph Network Environment

  Extract Secondary & Advanced Features
- Different Audience in different areas make various unique characteristic of the user categories
- Not being able to train the machine learning models as we can do it for other problems

https://twitter.com/samirsinh189

New York Art Beat
6,863 Tweets

Follow

New York Art Beat
@NYArtBeat

NY's most comprehensive art & design listings & reviews site. Daily updates for 700 events from 1100 galleries & museums. iPhone & Android app.

NYC nyartbeat.com Joined October 2008

2,650 Following **212K** Followers

Not followed by anyone you're following

Samir Dattopadhye
2,616 Tweets

Follow

Samir Dattopadhye
@samirsinh189

Proud Indian || Engineer by training || Dean @AniruddhasADM - a Disaster Management NGO, and other sister organisations ||
MD @LotusPubl'ications Pvt. Ltd.

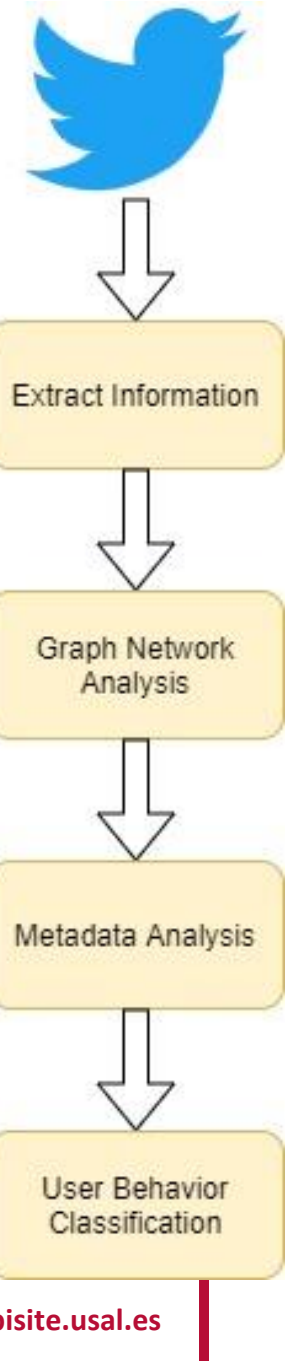
Mumbai, India aniruddhafriend-samirsinh.com Joined August 2015

556 Following **8,374** Followers

Not followed by anyone you're following

Influencer in Art VS **Influencer in Politics**

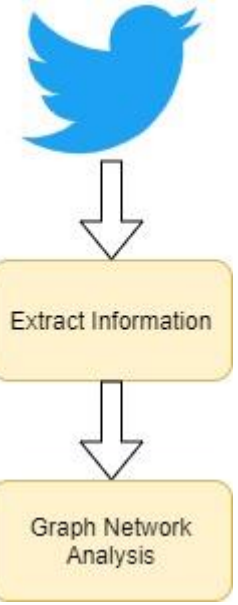
Proposed Architecture



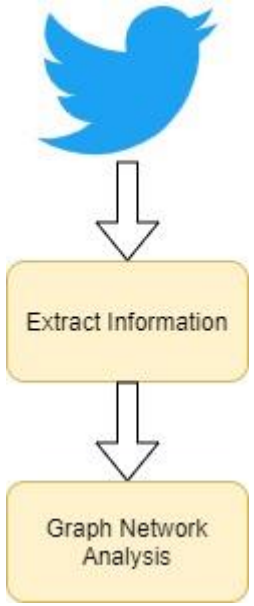
Graph Network Analysis

1. **Nodes** → Twitter Users
2. **Edges**
3. **Weights**
   - Retweets
   - Quotes
   - Mentions

Iteration of the connection between 2 Nodes



Graph Network Analysis



Centrality measures of the nodes:

- Eccentricity
- Closeness Centrality
- Betweenness Centrality
- Degree Centrality
- In-Degree Centrality
- Out-Degree Centrality

Topology measures:

- In-Degree Measure
- Out-Degree Measure
- Degree Measure

Function Measures & Algorithms:

- Has Self-Loops
- Community Detection Algorithms

Terms	Definitions
Eccentricity	The maximum shortest distance of one node from others. The lower the Eccentricity, the greater is the power of the node to influence others.
Clustering Coefficient Centrality	The nodes in a network that tend to be in the same cluster based on the degree of the nodes. $cc = \frac{n}{t}$
Closeness Centrality	Indicates how close a node is to the other nodes in a network by capturing the average distance based on one vertex to another. $cl = \frac{1}{\sum_{v \neq u} d(u,v)}$
Betweenness Centrality	Shows how influential the node is. The greater the value of betweenness centrality is, the more important that node would be to the shortest paths through the network. So, if that node is removed, many connections would be lost. $b = \sum_{s \neq v \neq t} \frac{\delta_{st}(u)}{\delta_{st}}$
Harmonic Closeness Centrality	This measure is so similar to closeness centrality, but it can be used in networks that are not connected to each other. This means that when two nodes are not connected, the distance will be infinity, and Harmonic Closeness is able to handle infinity just by replacing the average distance between the nodes with the harmonic mean.
In-Degree Centrality	This centrality indicates the importance via the number of edges entering the node.
Out-Degree Centrality	This centrality indicates the importance via the number of edges going out of the node.
Degree Centrality	This measures how many connections a node has. In other words, it is the summation of the In-Degree and Out-Degree of the node and shows how important a node is, in terms of the number of connections. Deg(v) = InDeg(v) + OutDeg(v)

Metadata Analysis

Basic Feature extracted from Tweet objects:

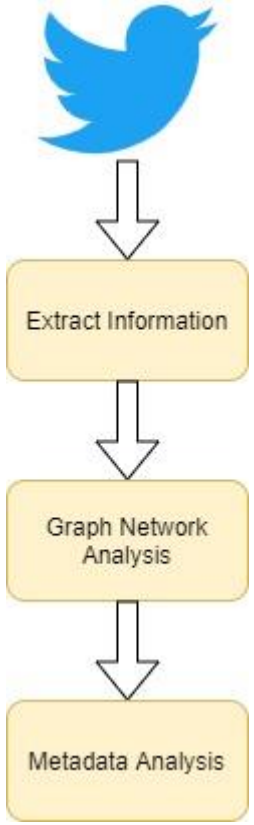
- User screen name and Id, etc.
- Information of the retrieved Tweet.
- Number of Likes, retweets and replies of the Tweet.
- Number of followers and followings of the user.
- And so on.

Secondary Features: (Doing a new Query for each profile)

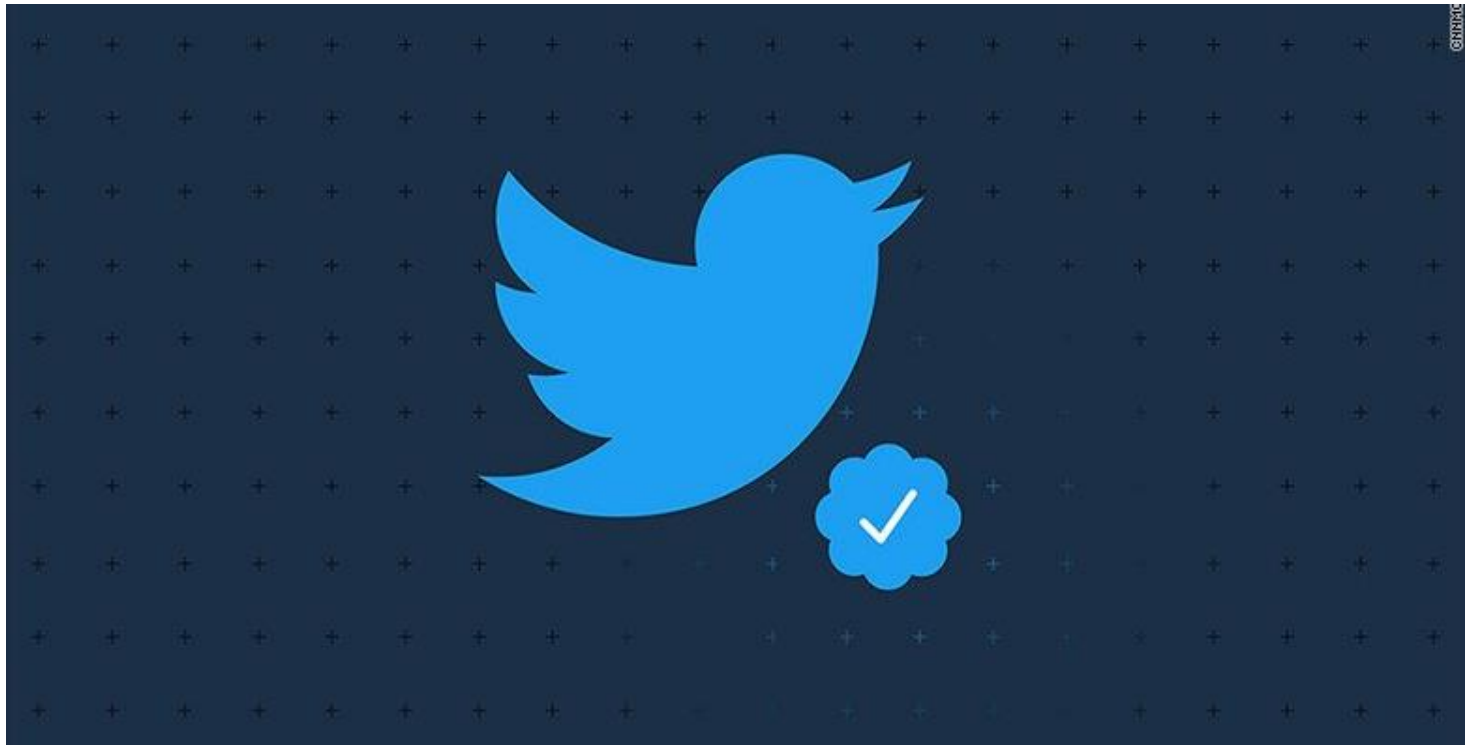
- Maximum tweets per day.
- Number of retweets.
- Number of tweets with URL.
- Etc.

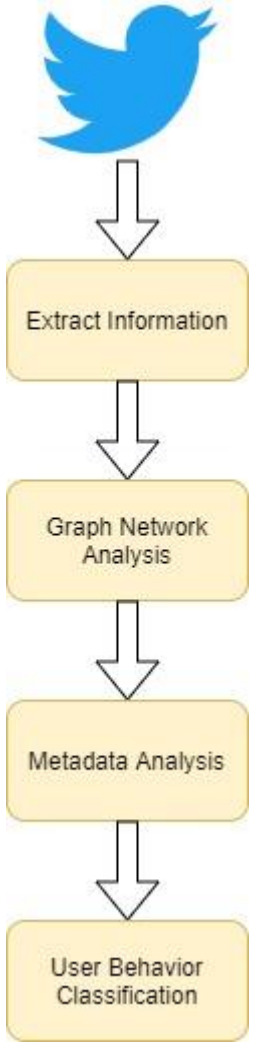
Create advanced features related to the previous ones

- Ratio of published tweets and RT.
- Ratio of tweets per year.
- Ratio of time gap between tweets.
- Etc.



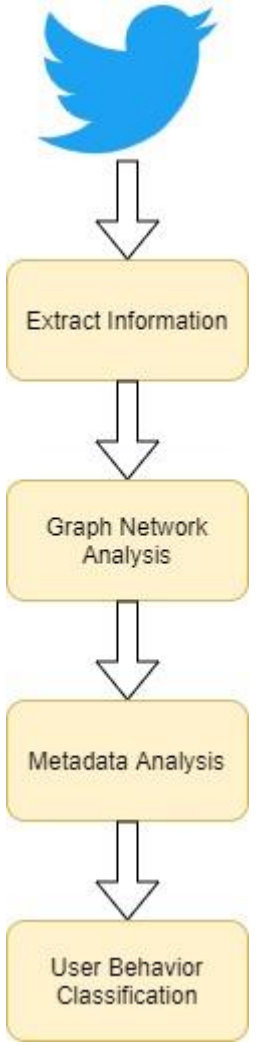
Verified Profiles





- Accounts which are officially verified by Twitter ⟶ *Verified = True*

Influencers

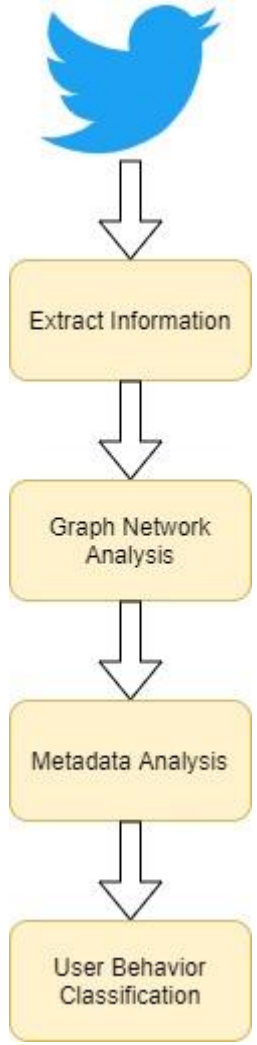


- People who are not verified by Twitter, but they have a great influence on others and their content has an undeniable engagement
  - A high number of their tweets
  - A high number of followers
  - The low time between tweets
  - The high number of interactions with their own tweets
  - High in-degree centrality

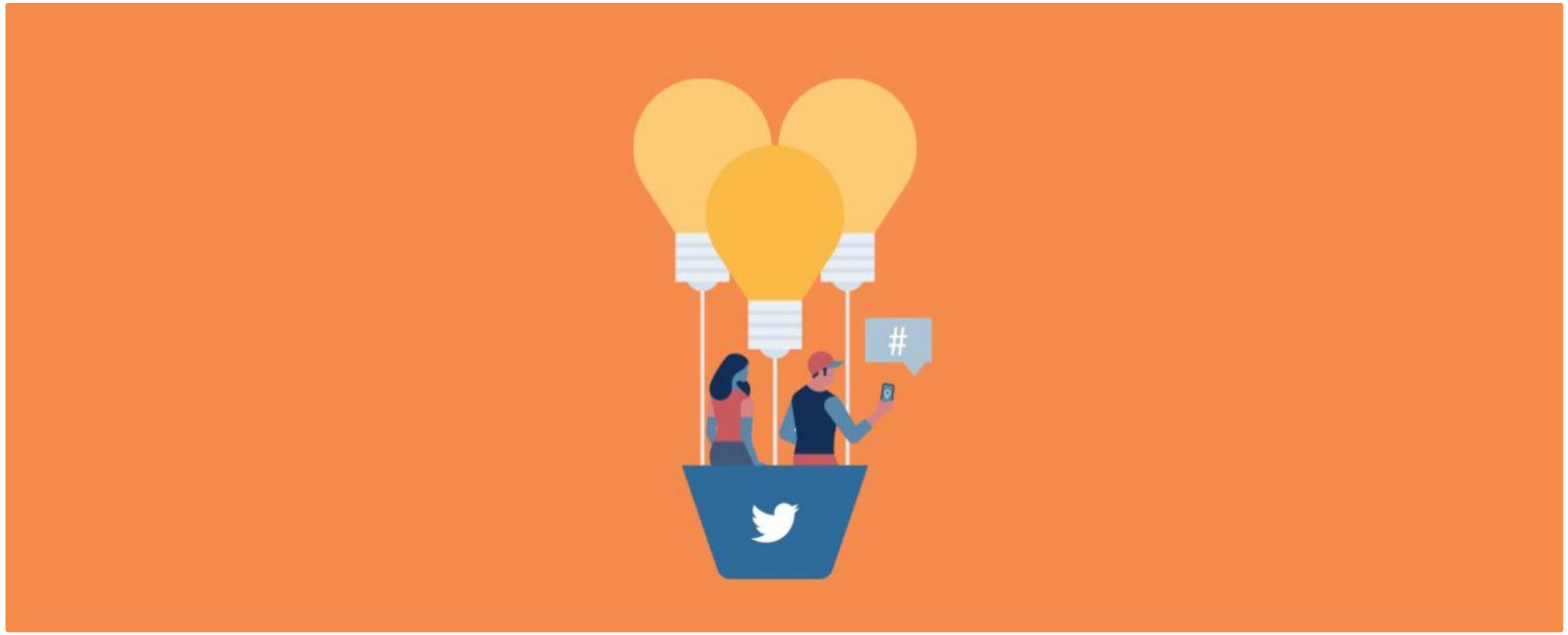
Fakes

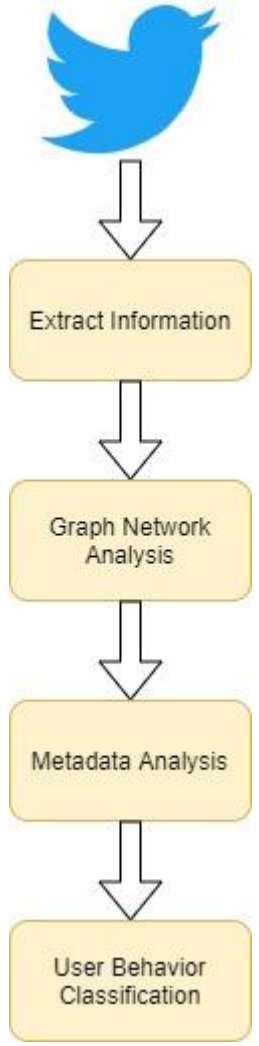
- People who are not verified by twitter, but their contents are fake news, spam, incoherent tweets, etc.

- The high number of Retweets
- The low time between tweets
- Default image profile
- No biography
- Numbers on its account name
- The small number of followers
- 2001 followings
- Tweets duplicated
- Self-Loop
- High outdegree number
- Low indegree number



Regular Profiles

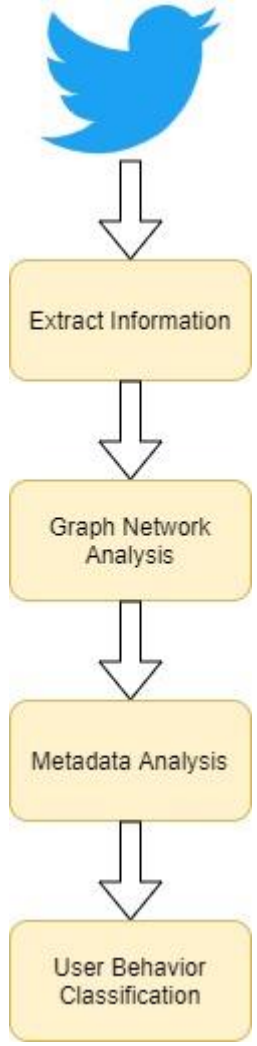




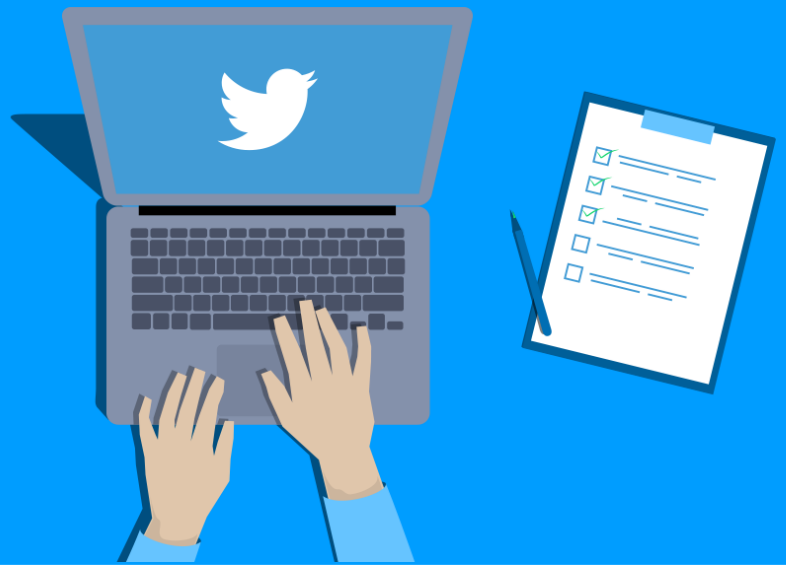
- People who are no verified by twitter, publish a few contents, favorites, with a balanced number of followers and followings, not in large numbers

User Behavior Classes

User Category	Characteristics	Related Features
Verified Accounts	Accounts which are officially verified by Twitter	Verified = True
Influencers	People who are not verified by Twitter, but they have a great influence on others and their content has an undeniable engagement	- A high number of their tweets - A high number of followers - The low time between tweets - The high number of interactions with their own tweets - High in-degree centrality
Regular Profiles	People who are no verified by twitter, publish a few contents, favorites, with a balanced number of followers and followings, not in large numbers	
Fakes	People who are not verified by twitter, but their contents are fake news, spam, incoherent tweets, etc.	- The high number of Retweets - The low time between tweets - Default image profile - No biography - Numbers on its account name - The small number of followers - 2001 followings - Tweets duplicated - Self-Loop - High outdegree number - Low indegree number

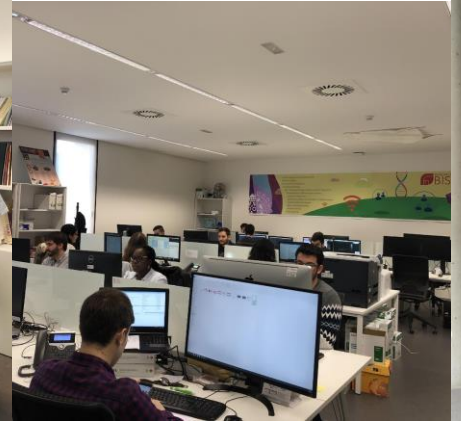


Futurework



- Add the Real-Time Mode to this tool.
- Integrate other Social Media Platforms.
- Use this user classification for further recommendations in different industries
- Generating reliable Datasets





Your Questions

Thank you!

Niloufar.Shoeibi@Usal.es

bisite@usal.es

fb.com/bisite

@bisite_usal

bisite_usal

bisite-research-group

Grupo de investigación BISITE

PyData SALAMANCA

bisite.usal.es