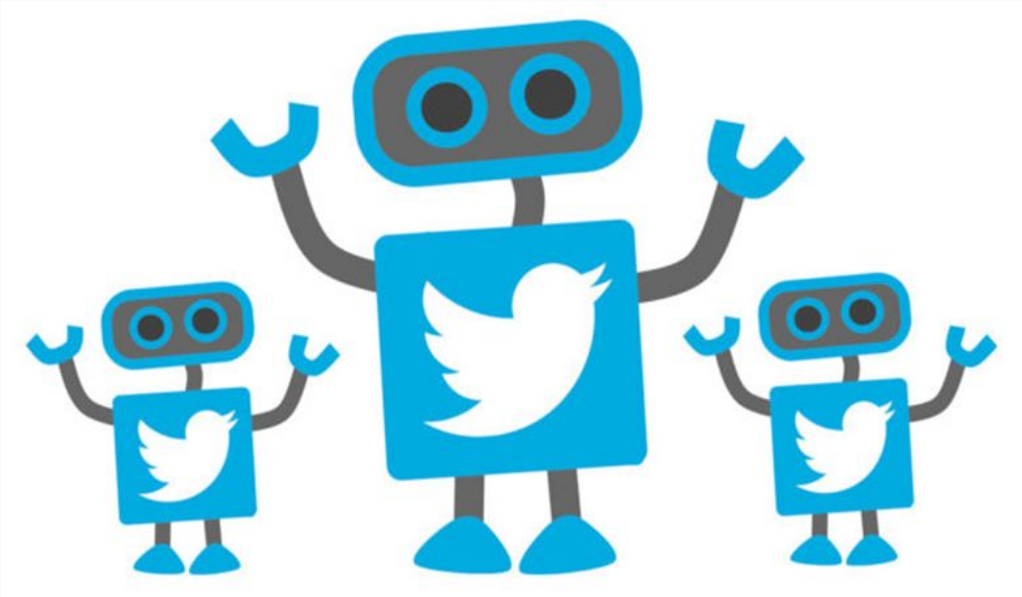
# AI APPLICATIONS: BOT DETECTION IN TWITTER



ICMAT
INSTITUTO DE CIENCIAS MATEMÁTICAS

**José Manuel Camacho Rodríguez**

**ESCUELA JAE DE MATEMÁTICAS 2021 -23th June 2021**

1. Introduction and motivations
2. Theoretical concepts and model comparison
3. Case study

# 1. Introduction and motivations

# What is a bot?

- A bot is a socio technical entity whose aim is to simulate human behaviour in Online Social Networks (OSN) .

- Bots resemblance as humans to other human users and to the OSN platform .

- Through different mechanisms, including Artificial Intelligence (AI), bots interpret the situation and react .

# Background and motivations (1/2)

- A bot in Twitter is regarded as a credible source of information.
- Bots are 2.5 times more influential than humans.
- Around 15% of accounts in Twitter are operated automatically or semi-automatically.
- More than 71 % of Twitter users mentioning US stock market are likely to be bots.

# Background and motivations (2/2)

- Bots are crucial in the propagation of fake news and misinformation in Social Networks. Impact in democratic processes:
  - *Brexit referendum*
  - *2017 French presidential election*
  - *2016 US presidential election*
  - *2014 Venezuelan protest*
- Malicious practices of bots in the stock market.

**Essential problem** which is faced by researchers and platforms.

# 2. Theoretical concepts and model comparison

# Bot detection problem

Classification problem: (Supervised machine learning problem)

$$f(x,\alpha) = \begin{cases} 1 & (Bot) \\ 0 & (\text{ Human }) \end{cases}$$

$x \equiv$ input vector of features

$\alpha \equiv$ hyperparameters

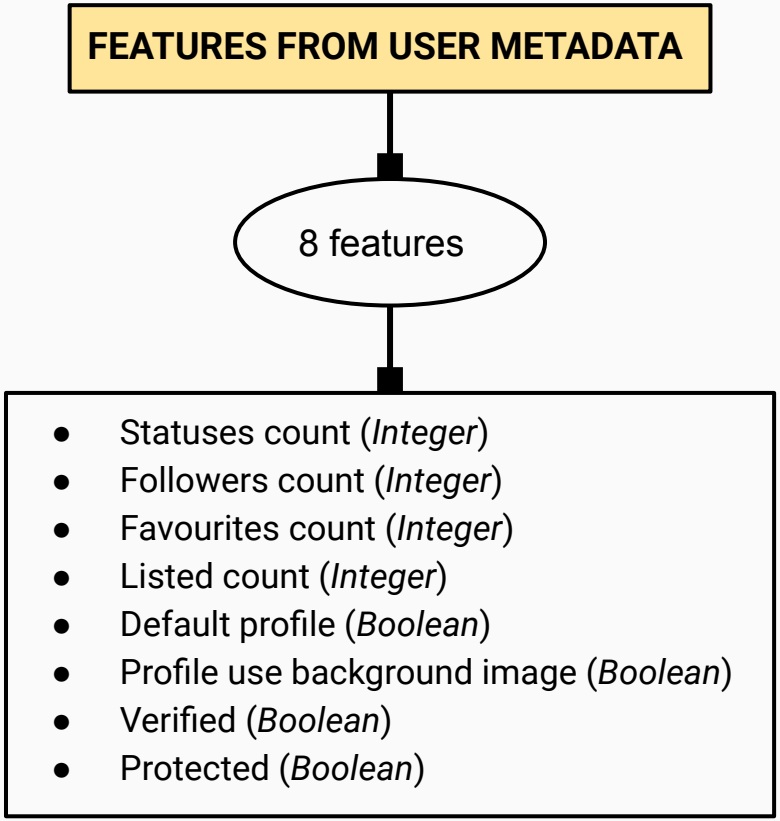
- Scalability vs performance
- Adversarial nature

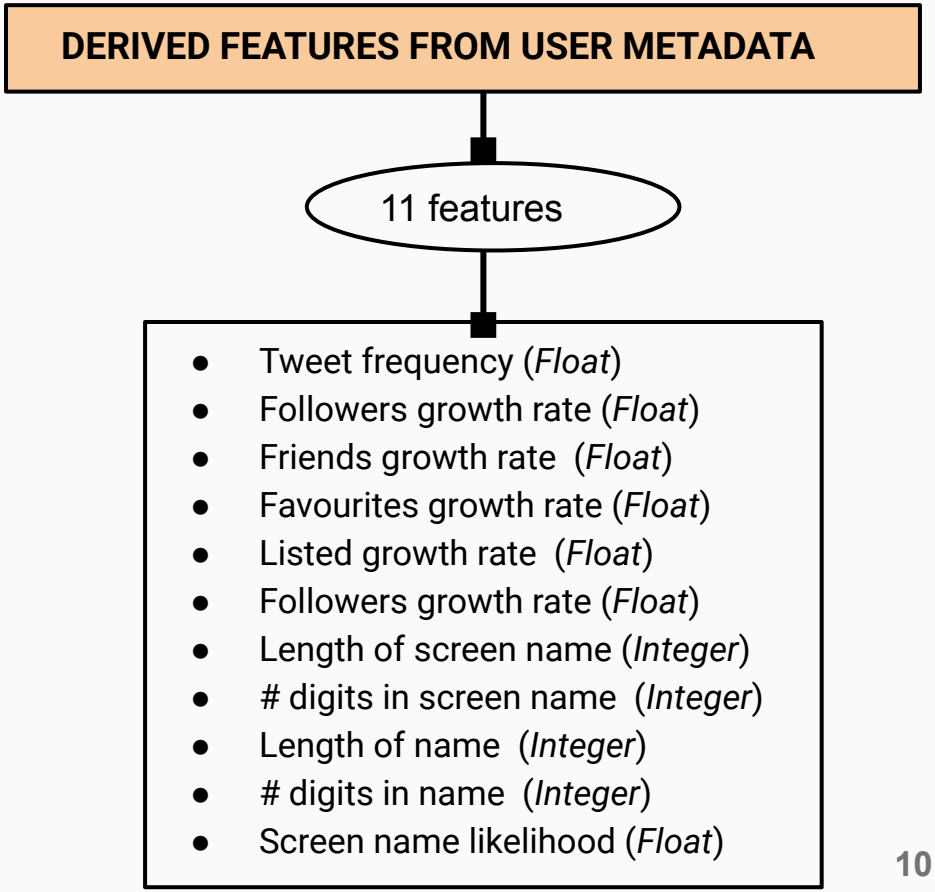
# Input features for bot detection model

- User metadata and derived features.
- Compression statistics from DNA string obtained from the Social Fingerprinting.

These features are used to feed a classification algorithm.

# Features from user metadata and derived features



# Social Fingerprinting: DNA string

- **Encode the behaviour of an account.**
- We use an **alphabet**.

$$\mathbb{B}^3_{\text{type}} = \left\{ \begin{array}{ll} A \leftarrow & \text{tweet} \\ \text{C} \leftarrow & \text{reply} \\ \text{T} \leftarrow & \text{retweet} \end{array} \right\} = \{\text{A}, \text{C}, \text{T}\}$$

Let's suppose @Twitter has the following timeline:

4:00 pm: *Retweet*, 4.23 pm :*Tweet*, 4:57 pm:*Tweet*, 5:02 pm: *Reply*,

Then, the DNA string of @Twitter is **TAAC.**

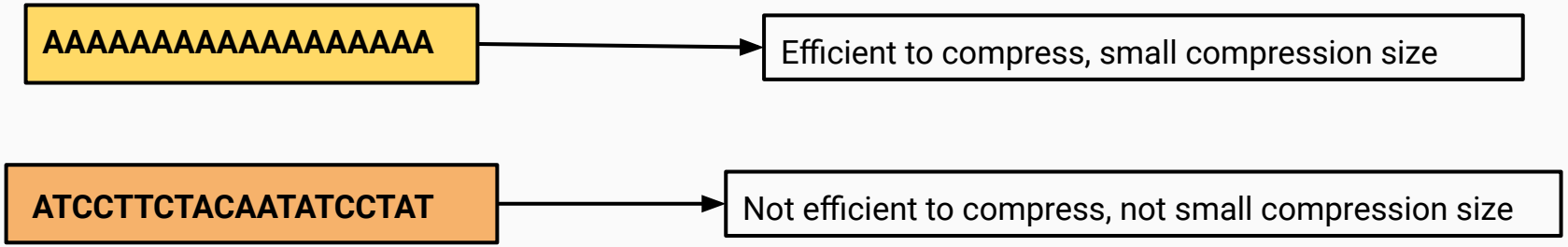
# Entropy and lossless compression algorithms

- **Entropy(Information theory)**: measure to evaluate the randomness of a data signal.

- **Lossless compression algorithm**: data compression algorithm which enables the perfect reconstruction of data from compressed data.
    - It is not efficient to compress and decompress data with a high entropy with a lossless compression technique.

# How to exploit DNA string: Use of compression algorithms

**Intuition:** Human-operated accounts display more random behaviour than bot-operated accounts

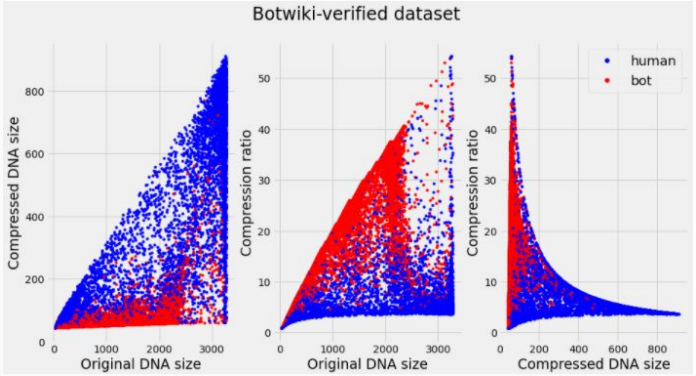


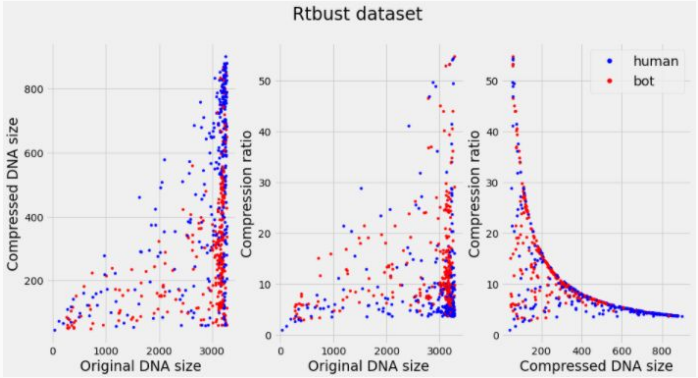

**Features extracted:**

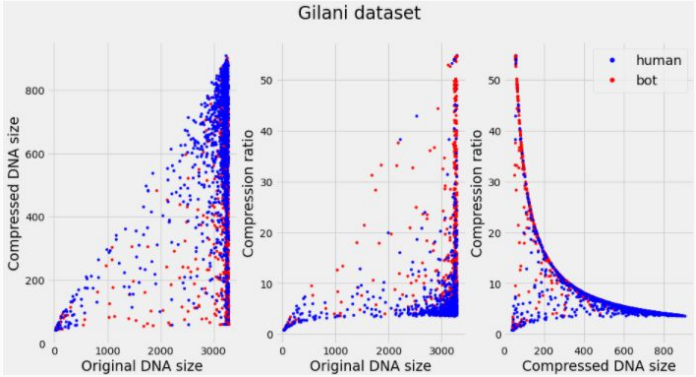
- Original DNA Size
- Compressed DNA Size
- Compression ratio

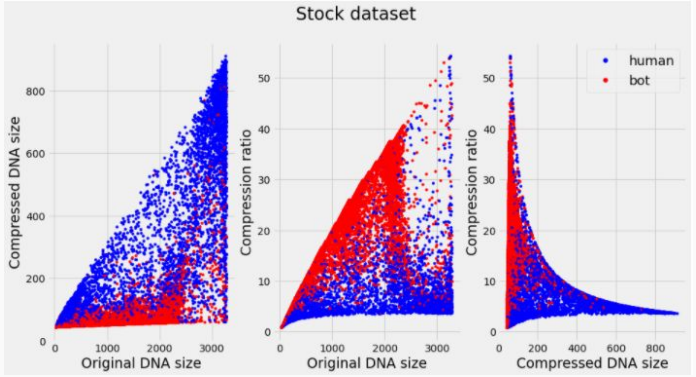
# Public datasets represented by compression statistics

Bot repository: *https://botometer.osome.iu.edu/bot-repository/datasets.html*



# Model comparison - procedure

- This model comparison is based on [Antenore et al. 2020] .

1. We consider the set of features:

   Light: 7 features from user metadata + 11 derived features
   DNA: Original DNA Size, Compressed DNA Size, Compression ratio

2. We use some of the datasets from the repository for training.

3. We feed different classification algorithms with the two previous set of features and training data.

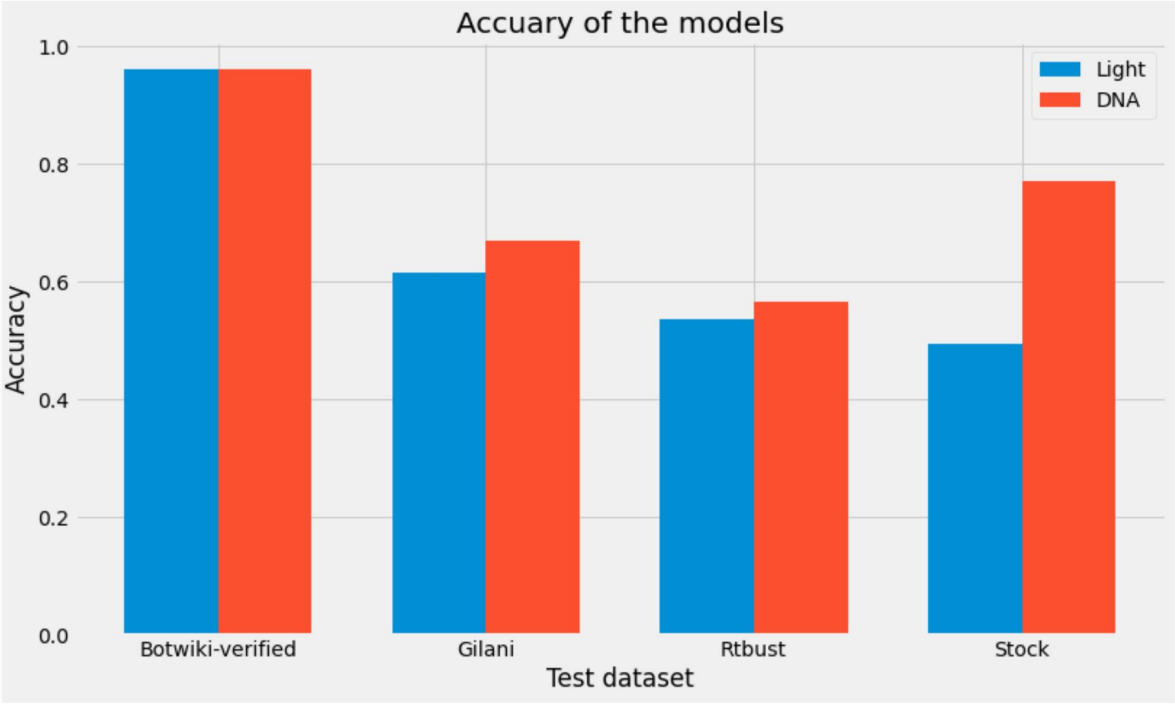
   *Random Forest , Logistic Regression , AdaBoost, Gradient Boosting, Linear SVM, K Nearest Neighbors, Gaussian Naive Bayes, Multilayer Perceptron ...*

4. Different datasets from those used from training are used to assess the performance of the model: *Botwiki-verified ,Gilani*, *Rtbust*, *Stock*. *(Test sets)*
   - *Metrics used: Accuracy*

# Model comparison - results

Light set: *Gradient Boosting*

DNA set: *Logistic Regression*



# Model comparison - conclusions

**DNA model** :

- Better performance
- Not scalable
- Cannot evaluate protected accounts or accounts with no timelines.

**Light model**:

- Good performance in most of the cases
    - Not perform well when bots display coordinated behaviour.
- Scalable
- Can evaluate protected accounts and accounts with no timeline

# 3. Case study

- We consider key topics of the pandemic during certain periods.

- We study:

    - Differences of proportion and activity of humans and bots

    - Differences on the discourse

        - Sentiment analysis

        - Hashtag analysis

# Topics and periods of the Covid-19 pandemic studied

**WUHAN -OUTBREAK** *(25th January -26th January)*:

- Virus spread in China
- Cancelation of large-scale events
- Travelling restrictions for more than 30 million people

**COVID** *(28th March-29th March)*:

- Trump considers quarantining New York
- Shortage of equipment for health workers and hospitals overloaded
- Milestone of 2000 deaths in US

**LOCKDOWN** *(10th May)*:

- First steps out of the UK lockdown

**TRUMP** *(4th February -21st February)*:

- State of Union
- Diamond Princess
- Problems with Covid-19 testing

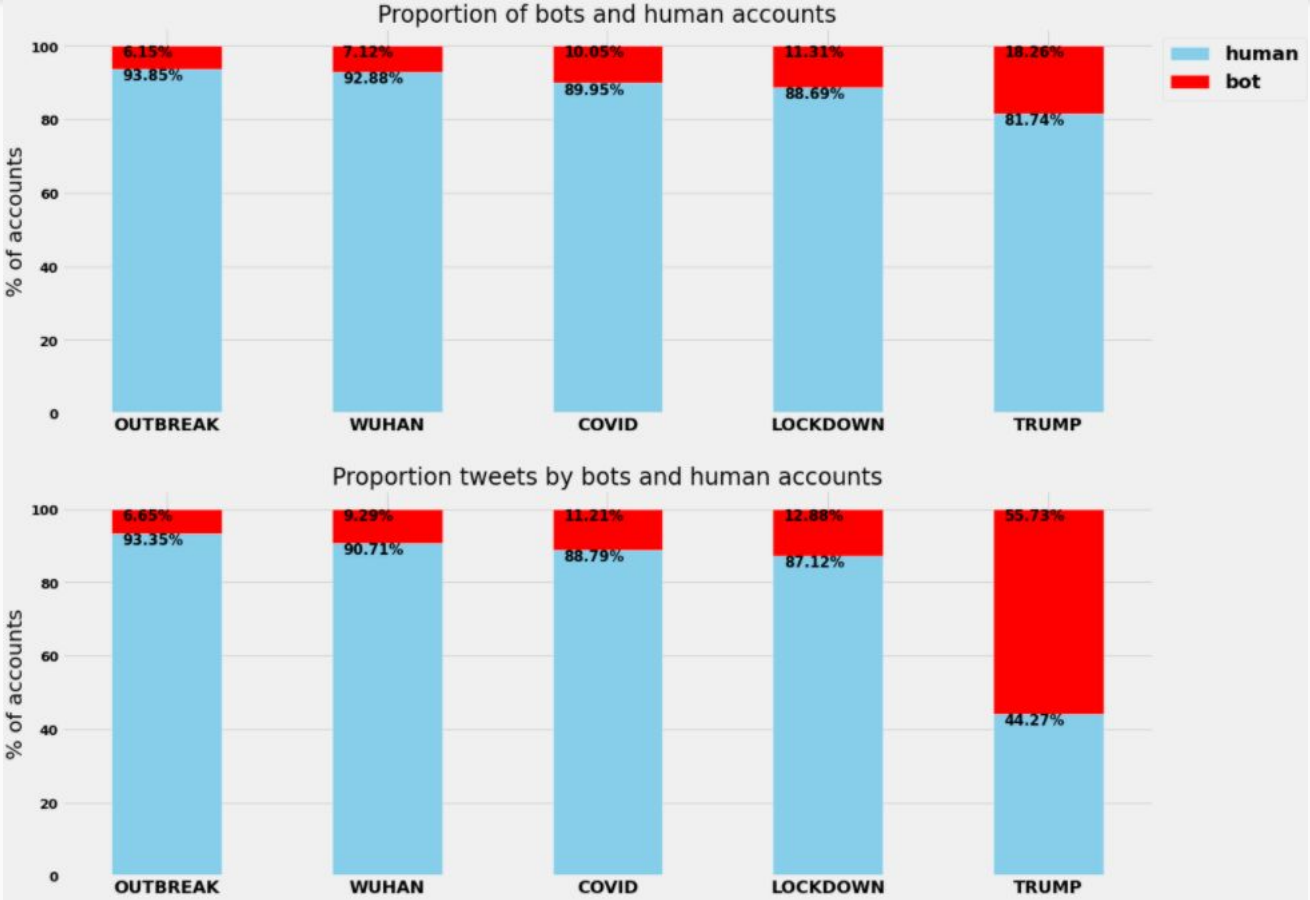
# Extraction of tweets for the experiment

- Topic: tweets with hashtags which contain all the variants of a substring.
  Ex: *substring*: covid ⟶ *hashtags:* #covid, #COVID19 ,#covid19, #COVID, #CoViD ...

| **Topic (substring)** | **Interval of time** | **# of accounts** | **# of tweets** |
|---|---|---|---|
| outbreak | 25/01/2020 - 26/01/2020 | 64602 | 82030 |
| wuhan | 25/01/2020 - 26/01/2020 | 103916 | 163723 |
| covid | 28/03/2020- 29/03/2020 | 312034 | 414097 |
| lockdown | 10/05/2020 | 26813 | 31052 |
| trump | 04/02/2020-21/02/2020 | 10144 | 26865 |
| | | **Total: 517509** | **Total: 717677** |

- We use **Light model**: scalability and good performance overall

# Proportion and activity of humans and bots



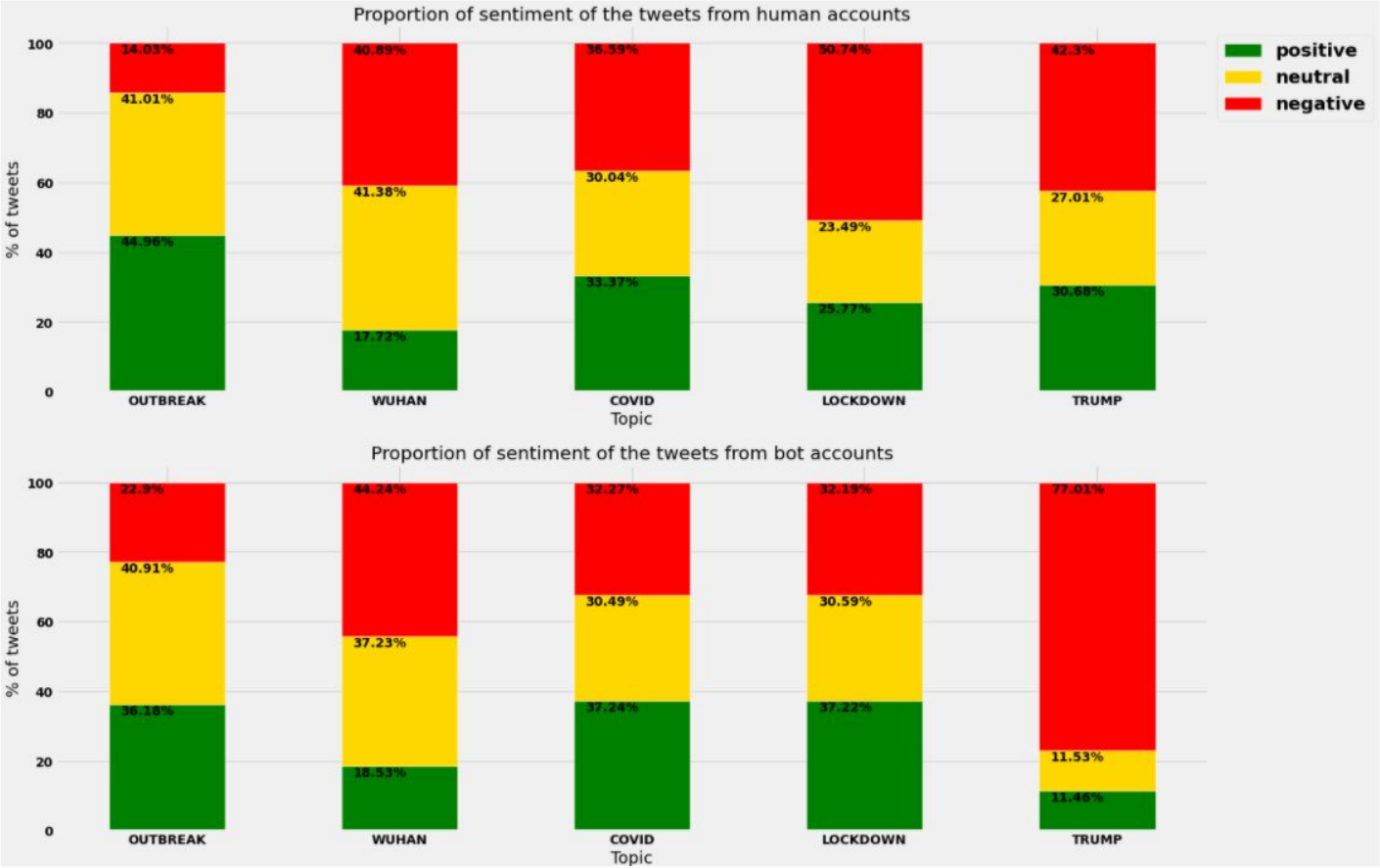
# Differences in the discourse : Sentiment analysis

Sentiment analysis : process of determining the attitudes, opinions and emotions expressed within a series of word.

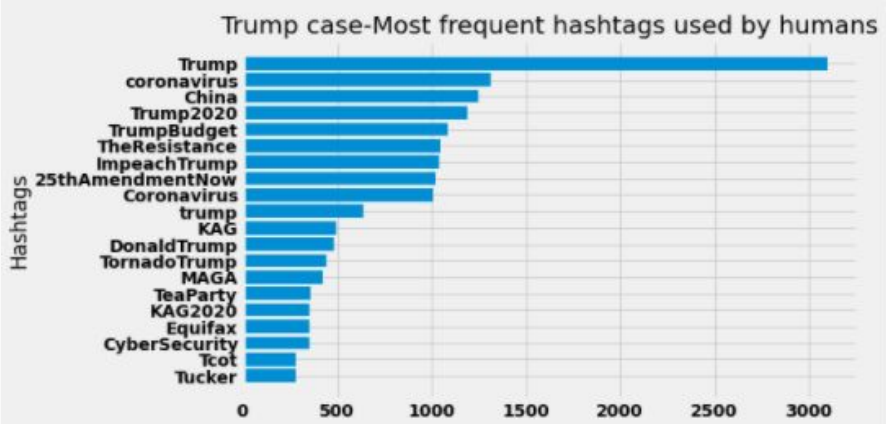
We consider it as a three class problem:

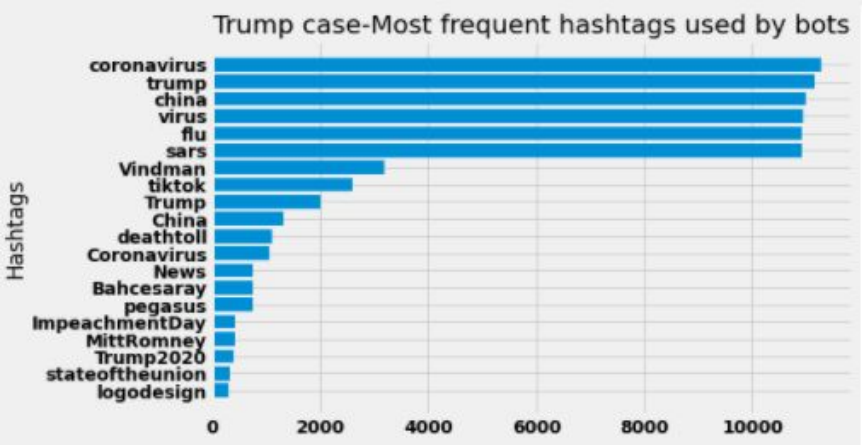
- Positive: *Trump is a good politician.*
- Negative: *Trump is a bad politician.*
- Neutral: *Trump is a politician.*

# Proportion of sentiment regarding bots and humans



# TRUMP hashtag analysis




Humans:

- pro-Trump hashtags: *#Trump2020,#KAG2020, #KAG, #MAGA, #TeaParty, #Tcot*
- Reference to budget proposal: *#TrumpBudget*
- Impeachment mentions attacking Trump: *#ImpeachTrump , #25AmendmentNow*

Bots:

- References to Covid-19: *#virus, #flu, #sars*
- Different way to mention impeachment: *#Vindman, #impeachmentDay,, #MittRodney*
- Also mentions State of union discourse: *#stateoftheunion*

# References

**[Antenore et al. 2020]** Marzia Antenore , Jose Manuel Camacho-Rodriguez , Emanuele Panizzi . **A comparative study of Bot Detection techniques methods with an application related to Covid-19 discourse on Twitter.** arXiv; 2021.

**[Antenore & Trinca. 2020]** Marzia Antenore and Elisabetta Trinca. **Who bots there, friend or foe? social bots and digital platforms. In Technological and digital risk.** Research issues. Peter Lang, 2020.

**[Yang et al.2020]** Kai-Cheng Yang, Onur Varol, Pik-Mai Hui, and Filippo Menczer. **Scalable and generalizable social bot detection through data selection.** In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 1096–1103, 2020.

**[Pasricha et al.2019]** Nivranshu Pasricha and Conor Hayes. **Detecting bot behaviour in social media using digital dna compression.** In 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science. AICS (Artificial Intelligence and Cognitive Science) 2019, 2019

*[**Sayyadiharikandeh et alt. 2020**] Sayyadiharikandeh, M., Varol, O., Yang, K. C., Flammini, A., & Menczer, F. (2020, October). **Detection of Novel Social Bots by Ensembles of Specialized Classifiers.** In Proceedings of the 29th ACM International Conference on Information & Knowledge Management (pp. 2725-2732).*

# THANK YOU FOR YOUR ATTENTION!