



# DO PEOPLE BELIEVE IN FAKE NEWS?

Agnès Gaspard  
Rémi Blaise  
Alan Collet

# Table of contents

## 1.Data Collection

## 2. Features

## 3. Used Method

- Naive Bayes
- LogisticRegression
- Simple Neural Networks

## 4. Results

- Accuracy/Matrix
- Demo

## 5. Problems and solutions

# Introduction

**FAKE  
NEWS**

## Context

- Lots of research on fake news detection
- Not one found on analysing readers' belief of the news

## Goal of project

Guess if a user believes or not to a news article via one of his comments

## Methods

NLP algorithms similar to sentiment analysis and machine learning classification methods

# Data Collection



POLITIFACT

kaggle



reddit

- Database with articles from Politifact
- Database with texts labelled with sarcasm
- Collect comments from Reddit

Labellized 3000 comments with

1: belief

-1: non-belief

0: none of the above

# Features

## Sarcasm

Method used instead of bag of words: Tf\*idf :

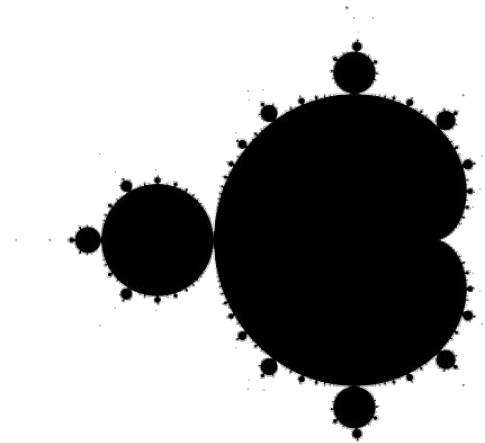
Term frequency

Inverse document frequency

## Sentiment Analysis

Word representation: bag of words

- Using Naïve Bayes
- Using TextBlob



TextBlob

# Used Methods

- Naïve Bayes

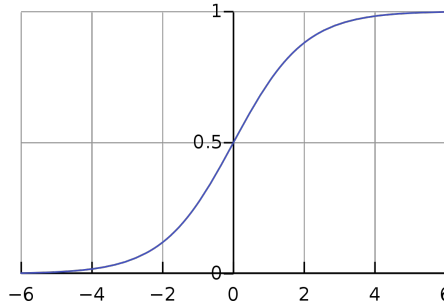
$$\log(P(Y = y, X)) = \log(P(X|Y = y)) + \log(P(Y = Y))$$

$$\log(P(Y = y)) = \log \frac{d\_count[y]}{d\_count}$$

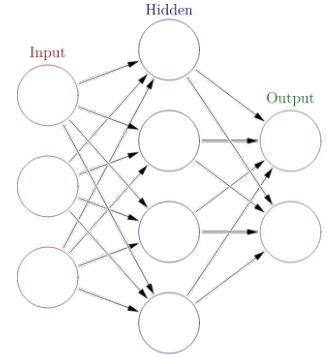
$$\log(P(X|Y = y)) = \sum_{w \in V} \log \frac{w\_count[w][y] + \alpha}{w\_count[y] + \alpha|V|}$$

$$P(Y = y|X) = \frac{P(Y = y, X)}{\sum_{y' \in Y} P(Y = y', X)}$$

- LogisticRegression



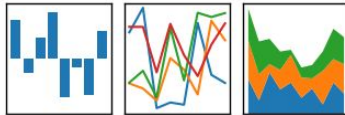
- Simple Neural Networks



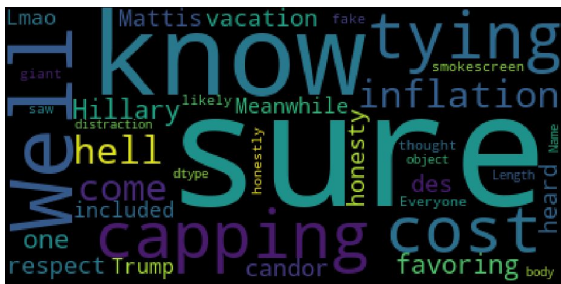
## Librairies used:

pandas

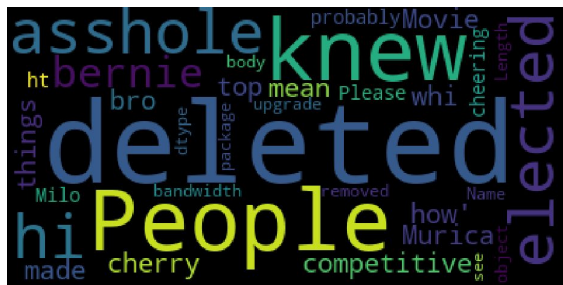
$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



# Results



## Class -1



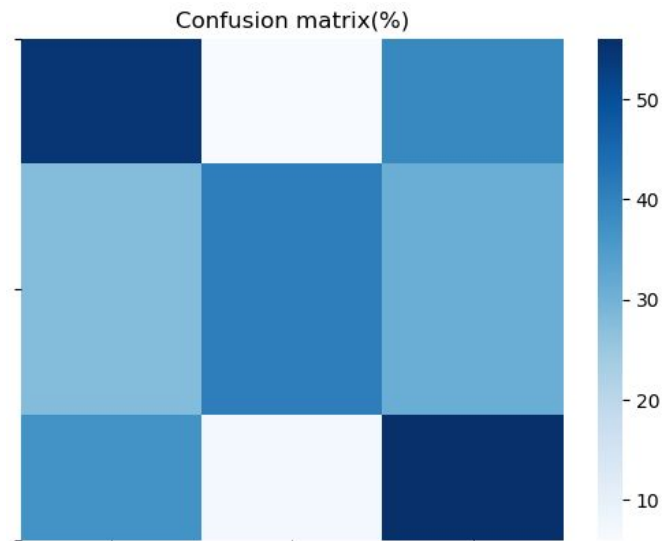
## Class 0



## Class 1

### Accuracy:

- Naive Bayes: 56% (Binary Class)
- Logistic Regression: 49% (3 Classes)
- Neural Networks: 46%(3 Classes)



# Problems encountered & solutions

Problems	Solutions
Lots of comments on one subject	Shuffle data before labellization
Lots of comments non related and fights	0 to not disturb training
Authors did not comments on their belief, mostly on articles subjects	No solution really found
Unbalanced data (204 “non-belief” comments out of 3000)	<ul style="list-style-type: none"><li>• SMOTE</li><li>• Labelled more data</li><li>• Reduce “belief” comments size</li><li>• Weight class</li></ul>



# Questions?