

Election Campaign Dynamics

Software Project - Final Presentation

Anahita Poulad

Emre Canbazer

Amir Jamali

Melina Chioutakou

February 11, 2022

OUTLINE

1. Introduction
2. Data Gathering
3. Topic Modeling
4. Topic Relation
5. Evaluation
6. Website
7. Further Improvements

Introduction

INTRODUCTION

- Motivations and Initial Ideas
 - Cognitive shortcuts for citizens ?
 - Voting Assistance Applications 😊
 - Political Compass Application 😊
 - Corpus Generator and Topic Modeler 😊
 - NLP-powered political science research

Data Gathering

DATA GATHERING



- Goal : Output a .csv corpus.

- easy to manipulate (pandas, Excel)
- tried .tsv - it was not very user-friendly



- What sources?

- News articles 😊 - legal issues and bias
- Social media 😊
 - Twitter - Good source, amazing API
 - YouTube - Lots of speech, powerful API offered by Google

DATA GATHERING: TWITTER DATA



- Real-time data extraction from candidates' Twitter accounts.
- using **tweepy** package
- Started with 7 features.
- Now 14 features (+2 for separated time features).

DATA GATHERING: TWITTER DATA

| | name | follower_count | tweet_text | fav_count | rt_count | time |
|------|--------------|----------------|--|-----------|----------|---------------------|
| 0 | Anasse Kazib | 46142 | Liberté pour les Ouïghours ! ❤️ https://t.co/Ut... | 375 | 52 | 2021-10-02 19:21:45 |
| 1 | Anasse Kazib | 46142 | Soutenu par @EricNauilleu et des Nazis https://... | 286 | 50 | 2021-10-02 15:28:34 |
| 2 | Anasse Kazib | 46142 | Et de 4 ce matin 🔥🔥🔥🔥\nMerci à ces maires de ... | 142 | 25 | 2021-10-02 15:22:50 |
| 3 | Anasse Kazib | 46142 | « Liberté pour Ouïghours » ❤️#Manifs2octobre #O... | 148 | 29 | 2021-10-02 13:16:49 |
| 4 | Anasse Kazib | 46142 | 🚫 Suppression de la figure présidentielle et d... | 43 | 7 | 2021-10-02 11:04:13 |
| ... | ... | ... | ... | ... | ... | ... |
| 1820 | Anasse Kazib | 46146 | Pour l'heure ici en France Macron reste le fav... | 37 | 7 | 2021-01-11 23:00:14 |
| 1821 | Anasse Kazib | 46146 | L'attitude autoritaire de Trump et Bolsonaro, ... | 49 | 9 | 2021-01-11 22:57:38 |
| 1822 | Anasse Kazib | 46146 | C'est interessant de voir comment le vent réac... | 67 | 13 | 2021-01-11 22:52:43 |
| 1823 | Anasse Kazib | 46146 | Lepen est tellement KO qu'elle est obligé de d... | 510 | 124 | 2021-01-11 22:47:09 |
| 1824 | Anasse Kazib | 46146 | Du coup on pourra pas tweeter le jour ou il y'... | 69 | 11 | 2021-01-09 23:56:33 |

1825 rows × 6 columns

[outdated] pandas dataframe created from the Tweets of Anasse Kazib

DATA GATHERING: TWITTER DATA

| | name | bio | follower_count | tweet_text | tweet_id | fav_count | rt_count |
|---|---------------------------------|---|----------------|---|---------------------|-----------|----------|
| 0 | Antoine WAECHTER 2022 🇩🇪🌐 | Candidat à la #Presidentielle2022 .\nMouvement... | 138 | 🇫🇷 La Charte 🇫🇷 du #MEI \n\nLa charte l les-ecol... | 1453649613722894336 | 2 | 1 |
| 1 | Antoine WAECHTER 2022 🇩🇪🌐 | Candidat à la #Presidentielle2022 .\nMouvement... | 138 | https://t.co/AAnlqz5GSF\n\nLa Conférence des N... | 1453648613721137153 | 2 | 1 |
| 2 | Antoine WAECHTER 2022 🇩🇪🌐 | Candidat à la #Presidentielle2022 .\nMouvement... | 138 | RT @DBarths: L'écologie est au cœur des débats... | 1453624625028648960 | 0 | 1 |
| 3 | Antoine WAECHTER 2022 🇩🇪🌐 | Candidat à la #Presidentielle2022 .\nMouvement... | 138 | RT @audreygarric: Le constat de l'@UNEP, à que... | 1453254441302429702 | 0 | 73 |
| 4 | Antoine WAECHTER 2022 🇩🇪🌐 | Candidat à la #Presidentielle2022 .\nMouvement... | 138 | https://t.co/0YpASOO9oO \n🇫🇷 TRIBUNE 🇫🇷\n\nCO2 : ... | 1452829334939619330 | 0 | 0 |

[updated] pandas dataframe created from the Tweets of Antoine Waechter

DATA GATHERING: TWITTER DATA

| rt_count | in_reply_to | mentions | is_retweet | is_quote | hashtags | emojicons | time |
|----------|-------------|--------------------|------------|----------|--------------|---|---------------------|
| 1 | None | None | False | False | MEI | :red_circle:, :scroll: | 2021-10-28 09:07:30 |
| 1 | None | COP26, cop21_paris | False | False | waechter2022 | :United_Nations:, :leaf_fluttering_in_wind:, :... | 2021-10-28 09:03:31 |
| 1 | None | DBarths | True | False | None | None | 2021-10-28 07:28:12 |
| 73 | None | audreygarric, UNEP | True | False | COP26 | None | 2021-10-27 06:57:13 |
| 0 | None | None | False | False | waechter2022 | :red_circle:, :newspaper:, :chart_increasing:,... | 2021-10-26 02:48:00 |









[updated] pandas dataframe created from the Tweets of Antoine Waechter

DATA GATHERING: YOUTUBE DATA



- Transcriptions from political speeches.
- using **googleapiclient** (metadata) and **youtube-transcript-api** (captions).
- 8 features
- Tried human-transcribed captions, but chose automatically generated ones.

DATA GATHERING: YOUTUBE DATA

| video_title | video_id | video_description | upload_date | duration | time_caption | transcript |
|------------------------------------|-------------|---|---------------------|----------|-----------------|---|
| Discours de Marine Le Pen à Fréjus | KbvDa-sHATk |  http://www.adhesions-rn.fr  http://www.m... | 2021-09-13 08:33:58 | 39:56 | 0 days 00:35:04 | contemporains nous devons notre liberté à des ... |
| Discours de Marine Le Pen à Fréjus | KbvDa-sHATk |  http://www.adhesions-rn.fr  http://www.m... | 2021-09-13 08:33:58 | 39:56 | 0 days 00:36:04 | les défenseurs d'une france libre et du monde ... |
| Discours de Marine Le Pen à Fréjus | KbvDa-sHATk |  http://www.adhesions-rn.fr  http://www.m... | 2021-09-13 08:33:58 | 39:56 | 0 days 00:37:04 | beaucoup progressé et nous sommes devenus un p... |
| Discours de Marine Le Pen à Fréjus | KbvDa-sHATk |  http://www.adhesions-rn.fr  http://www.m... | 2021-09-13 08:33:58 | 39:56 | 0 days 00:38:04 | direction du rassemblement national j'y vois l... |

pandas dataframe created from the video of Le Pen's speech in Fréjus

Topic Modeling

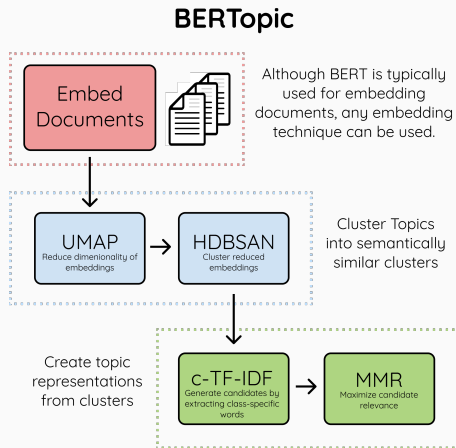
TOPIC MODELING

- **Key idea:** Analyze the political sphere
- **Approach:** Get meaningful information about the presidential campaigns and their inner structure

⇒ Topic Modeling with BERTopic on Twitter data

TOPIC MODELING WITH BERTOPIC

- Created by Grootendorst in 2020
- Topic modeling technique
- Uses BERT embeddings and class-based TF-IDF to create dense, informative clusters



TOPIC MODELING WITH BERTOPIC

- **Its goal:**
 - Provide easily interpretable topics
 - Keep important words in the topic descriptions
- **Additional advantages:**
 - Easy-to-use and accessible
 - Interactive visualization methods
 - Transparency of the results

KEYBERT AS AN ADDITIONAL STEP

- Keyword extraction algorithm
- From the creator of BERTopic
- Uses BERT embeddings to generate keywords and key phrases from a document based on their similarity to the document

⇒ retrieve more meaningful topics and minimize time

KEYBERT AS AN ADDITIONAL STEP

- For the creation of a BERTopic object:
 - Preprocessed list of tweets
 - Time and date
- Keyword extraction is performed after preprocessing on the tweets
- BERTopic is then built based on resulting list of keywords

PREPROCESSING

- Remove:
 - re-tweets
 - http/bitly links
 - hashtags
 - punctuation, double spacing
 - numbers
 - photos
 - emoticons
 - stopwords
- Lemmatization and removal of Verbs

MOST FREQUENT TOPICS

- Possibility to remove extra stopwords ⇒ remove words that do not provide information
- Gain deeper knowledge about the consistency of the topics
- e.g. access the keywords that comprise the topics/find most representative text of a topic

Global Topic Representation

- 0_français_france_être_père
- 5_ben_reportage_bravo_charlin
- 1_élection_candidat_politique_démocratie
- 2_youtube_chaine_intégralité_vidéo
- 6_musulman_islam_islamique_islamisation
- 7_merci_accueil_chaleureux_échange
- 8_europe_européen_emmanuel_macron
- 3_france_français_politique_campagne
- 9_ruralité_rural_agriculteur_paysan
- 4_école_enfant_parent_élève

Figure 1: Zemmour's most frequent topics from 01/09/2021 to 05/02/2022

VISUALIZATION TOPICS OVER TIME

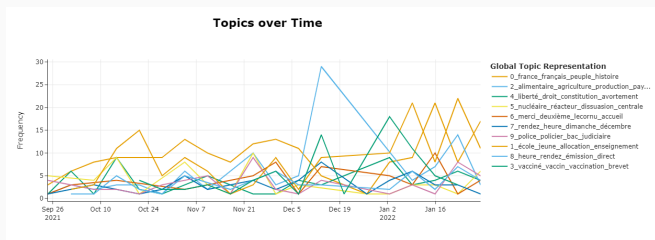


Figure 2: Mélenchon's 10 most frequent topics

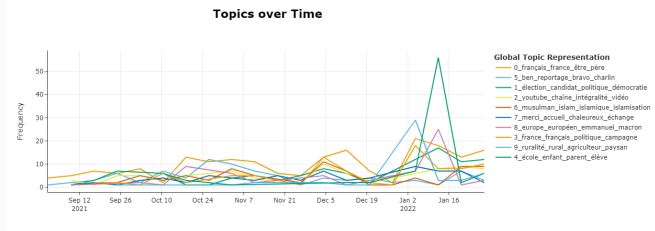


Figure 3: Zemmour's 10 most frequent topics

Topic Relation

TOPIC RELATION

- How do candidates' key points interact with each other?
- Important information in order to examine the political campaign atmosphere

⇒ SpaCy's *similarity* method to compute the semantic similarity between topics

SIMILARITY BETWEEN TOPICS

Similarity scores worth mentioning between the two candidates:

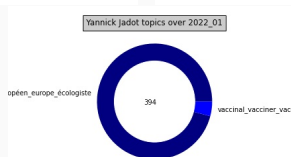
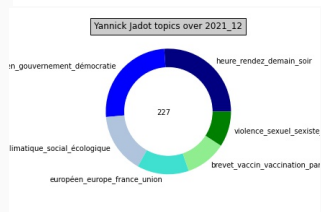
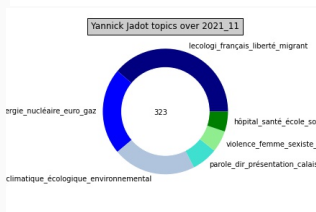
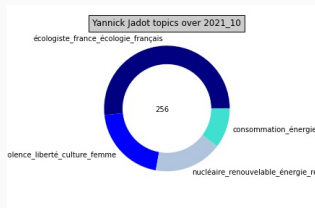
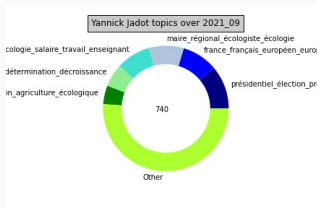
- "0_france_français_peuple_histoire" and "3_france_français_politique_campagne" \Rightarrow 0.84
- "1_école_jeune_allocation_enseignement" and "4_école_enfant_parent_élève" \Rightarrow 0.78
- "2_alimentaire_agriculture_production_pay" and "9_ruralité_ruralagriculteur_paysan" \Rightarrow 0.37
- "4_liberté_droit_constitution_avortement" and "1_élection_candidat_politique_démocratie" \Rightarrow 0.71

Drawback: Similarity scores are a mix of different signals and vectors trained on different data can produce very different results

SIMILARITY BETWEEN TOPICS

- "4_liberté_droit_constitution_avortement" and "1_élection_candidat_politique_démocratie" \Rightarrow 0.71
- Topic 4 words: ["liberté", "droit", "constitution", "avortement", "genre", "ivg", "fondamental", "république", "autonomie", "dignité"]
- Topic 1 words: ["élection", "candidat", "politique", "démocratie", "politicien", "présidentiel", "droite", "électeur", "idée", "autre"]

TOPIC DEVELOPMENT



TOPIC COMPARISON



Evaluation

EVALUATION

- **Challenges:**

- Topic Modeling evaluation
- Lack of expertise in the domain

- **3 Main Approaches:**

- Topic interpretability
- Topic coverage and diversity
- Topic coherence \Rightarrow NPMI
 \Rightarrow Adapted the implementation wrt BERTopic model

Website

TECHNOLOGIES

- Language: Python
- Framework: Flask
- Frontend: Html, CSS, JS
- Docker



DEMO

Election

Create Election Corpus:

Select one Candidate: ▼

From
2022-02-04 📅

To
2022-02-11 📅

Download

Show

@ Youtube Video Url

Download

Link to France2022

DEPLOYMENT

- Why Azure?
- Resource Group
- Azure Container Registry
- App Service
- Challenges and package conflicts!



Further Improvements

IMPROVEMENT AND BUG FIXES

- Improve fetching data from Twitter
- Build a pipeline for YouTube captions that checks the spelling, coherence and cohesion
- Create better graphical representations to present the importance of topics over time
- A better evaluation with OCTIS that can compare topic models by different measures in Classification metrics, Coherence, Diversity, Similarity, Topic significance

Thank you!