

Election Campaign Dynamics

Data gathering, Pre-processing, Topic modeling, Timeline

Anahita Poulad

Emre Canbazer

Amir Jamali

Melina Chioutakou

October 12, 2021

OUTLINE

1. Data Gathering
2. Preprocessing
3. Topic Modeling
4. Project Milestones and Schedule
5. Future Possible Approaches

Data Gathering

TWITTER

◇ From Twitter

Scrape tweets and store the information in lists

```
name = []
bio = []
follower_count = []
tweets = []
fav_count = []
rt_count = []
time = []

for i in tweepy.Cursor(api.user_timeline, id = candidate_username , tweet_mode = 'extended', since= date).items():
    if i.full_text[:4] != 'RT @': #excludes retweets
        name.append(i.user.name)
        tweets.append(i.full_text)
        time.append(i.created_at)
        bio.append(i.user.description)
        follower_count.append(i.user.followers_count)
        fav_count.append(i.favorite_count)
        rt_count.append(i.retweet_count)

# Creates a cursor objects and iterates through it to store tweet text and metadata of the tweet in lists
```

script to extract tweets of a candidate starting from a certain date

TWITTER

◇ From Twitter

	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 @EricNaulleau 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

pandas dataframe created from the lists

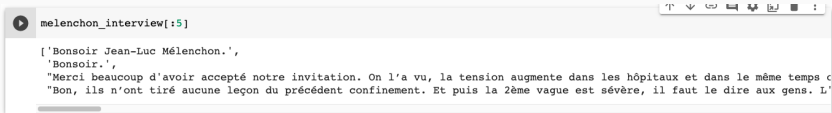
YOUTUBE

◇ From YouTube

```
video_id = "xxxxxx"  
# replace by the video id which is the part of the URL after 'https://www.youtube.com/watch?v='  
  
srt = YouTubeTranscriptApi.get_transcript(video_id, languages = ['fr'])  
# get and store the subtitle as a dictionary with metadata
```

script to extract YT captions using yt-transcript-api package

→ after cleaning



```
melenchon_interview[:5]  
  
['Bonsoir Jean-Luc Mélenchon.',  
 'Bonsoir.',  
 "Merci beaucoup d'avoir accepté notre invitation. On l'a vu, la tension augmente dans les hôpitaux et dans le même temps c  
 "Bon, ils n'ont tiré aucune leçon du précédent confinement. Et puis la 2ème vague est sévère, il faut le dire aux gens. L'
```

list containing the text of the interview split by speaker

DATA

Bonsoir. Bon, ils n'ont tiré aucune leçon du précédent confinement. Et puis la 2ème vague est sévère, il faut le dire aux gens. L'épidémie, c'est grave. Ce n'est pas une petite maladie qu'on a là, donc respectez les gestes barrières. Mais après ça, on voit bien que c'est la pagaille parce que l'incohérence des ordres et des contre-ordres a abouti à ce résultat que l'on ne respecte qu'un peu les consignes, ce n'est pas très grave. Et puis il y a des situations où on est arrivés au bout de la vie qui, que ce soit pour le commerce de proximité, les pratiques culturelles, l'école où ça semble être la pagaille partout. Il y a beaucoup de souffrance et je crois que malheureusement, comme on nous annonce une 3ème et même une 4ème vague, il faut qu'on commence à réfléchir, mais cette fois-ci qu'on réfléchisse avant et sérieusement à des alternatives au confinement. Écoutez, là on vient de voter l'état d'urgence sanitaire, il y en a pour 6 mois. On ne va pas laisser la France confinée pendant des années, parce que notre société va s'effondrer si on met la France en prison pendant des mois. Donc il faut toujours prévoir, planifier, organiser à l'avance, pas attendre la dernière minute, et surtout arrêter de croire que les choses vont se faire toutes seules. Je crains que c'est ce qu'a fait le gouvernement : il n'a pas planifié. Voilà, c'est tout. Et donc c'est la pagaille. Alors manifestement, vous pensez que vous pourriez faire mieux. Écoutez, quand tout va mal et que ça semble être nuit noire pour beaucoup de gens qui ne trouvent pas leur compte dans la brutalité de cette société, il faut allumer une lumière pour qu'on se dise : il y a un bout au tunnel, on peut faire autrement. Et mon intention est d'aider à déconfiner les esprits, à nouveau à se projeter sur l'avenir. Alors, je ne dis pas que je sois une projection sympathique pour tout le monde, mais pour des millions de braves gens, simples, oui, c'est le cas et je les appelle à se regrouper. Alors oui, je suis prêt et je propose ma candidature. Mais, il y a un mais : je demande une investiture populaire. Par conséquent, je serais candidat définitivement si et seulement si j'ai recueilli 150 000 signatures de parrainage. À ce moment-là, je me sentirais investi par le peuple. Et bien sûr, ce n'est pas un homme qu'on investit seulement, ça n'aurait pas de sens. C'est un programme, c'est l'Avenir en commun, tout le monde le connaît. Je ne veux pas le résumer maintenant, je peux vous dire sa philosophie : l'harmonie entre les êtres humains et avec la nature. Moi je veux qu'on recommence à parler des conditions de vie sociales, de l'écosystème, des gens, comment ils vivent, qu'est-ce qu'ils attendent, leur dignité. Au lieu de passer notre temps à aller disputer sur la religion ou à essayer de les mettre au pas « metro, auto, boulot, dodo » . Non, mais j'ai posé la question parce qu'une députée, Madame Caroline Fiat, m'a dit « pose la question : « est-ce qu'il y a d'autres candidats ? ». Personne ne s'est proposé. Non mais il y a beaucoup de talents. Attention, non mais je préfère inquiéter ceux qui ne nous aident pas. S'ils en avaient fini avec moi, il y en aura d'autres. Il y a beaucoup de talents. J'ai d'ailleurs une équipe qui est prête à gouverner. On a un programme, une équipe prête à gouverner. Tout ça est un point d'appui qui peut être rassurant. Et puis bienvenue à tous ceux qui veulent participer à cet engagement. Vous savez, 2022 c'est quand même le moment de tourner, de changer. Il y a plein de gens de toutes opinions, pas seulement les gens de ma famille politique, qui voient bien que la société est en impasse. Que ses valeurs de pacotille ne valent rien, qu'il faut qu'on tourne. Alors 2022, c'est une occasion. Nous sommes un grand pays, nous avons les moyens d'innover, de faire les choses différemment. Et puis, surtout, d'abolir la monarchie présidentielle. Mais ça, il n'y a que moi qui le propose. J'ai l'impression que les autres se verraient bien en monarchie. Justement, vous parliez de la monarchie présidentielle, est-ce que vous avez l'intention de faire l'union à gauche ? Est-ce que pour être élu, la prochaine fois, vous avez besoin des écologistes notamment, et est-ce que votre candidature se soir ne ferme pas un peu la porte à tout ça ? Vous savez, d'abord j'ai pris le temps de regarder. Toutes les organisations et les partis politiques ont décidé, tous, d'avoir un candidat. Certains ont déjà plusieurs précandidats, donc les mois qui vont venir devant nous, ça va être plutôt des mois de pagaille. Et puis c'est du temps perdu, parce que c'est maintenant qu'il faut construire ce que moi je voudrais construire : une majorité pour. Vous comprenez, il y en a marre d'aller aux élections pour voter pour le moins pire. Il s'agit donc de construire quelque chose de positif, de commun, plutôt que des animosités. Alors bon, vous ne parlez de la gauche, je vous dis que tout le monde a annoncé qu'il y aurait un candidat. Parfait. J'ai compris, et du coup, je m'avance. Je suis un pôle de stabilité et puis on verra chemin faisant, qui vivra, verra. Ce que je peux dire, c'est que clairement la dispersion peut nous nuire, mais la dernière fois déjà c'était là. Mais je vais vous dire, pire que tout serait la confusion : des gens qui ne seraient d'accord sur à peu près rien, mais qui donneraient l'impression, comme ça, de se mettre d'accord juste pour montrer du doigt les autres. Si je recueille une investiture populaire de 150 000 membres. Alors tout le monde peut aller, dès maintenant, là, après m'avoir entendu, si ça vous intéresse, sur la plateforme internet qui s'appelle le nousmesseur.fr parce que « nous sommes pour » c'est le positif et c'est ça que je veux défendre, vous pouvez le faire dès maintenant. C'est moi qui vous remercie de m'avoir donné cette occasion de m'exprimer devant autant de monde. Elles sont prévisibles.

◇ From YouTube

← captions to .txt

← 6-minute video interview

← only the candidate's speech

← 1128 tokens

.TXT file generated from the captions

Preprocessing

PREPROCESSING

1. Basic cleaning

Remove:

- re-tweets
- http/bitly links
- hashtags
- punctuation, double spacing
- numbers
- photos
- emoticons

2. Basic cleaning + tokenization + removal of stopwords

PREPROCESSING

	name	tweet_text	preprocessed_tweet	basic_cleaning
0	Marine Le Pen	Les annonces du Premier ministre en matière de pouvoir d'achat sont un cautére sur une jambe de bois. Les Français n'ont pas besoin de mesurerettes électorales, mais d'une baisse réelle et durable du prix des énergies et de l'essence ! nMLP https://t.co/D3jFEKXcqZ	annonces premier ministre matière pouvoir achat cautére jambe bois français besoin mesurerettes électorales baisse réelle durable prix énergies essence mlp	les annonces du premier ministre en matière de pouvoir d achat sont un cautére sur une jambe de bois les français n ont pas besoin de mesurerettes électorales mais d une baisse réelle et durable du prix des énergies et de l essence mlp
1	Marine Le Pen	Très heureuse de poursuivre mes rencontres avec les Français, aux côtés de femmes et d'hommes de terrain à l'image de @diaz_edwige et @JPGarraud. #Cavignac #MLaFrance n n https://t.co/b7AyimklpR https://t.co/66tgKsZbtU	très heureuse poursuivre rencontres français côtés femmes hommes terrain image	très heureuse de poursuivre mes rencontres avec les français aux côtés de femmes et d hommes de terrain à l image de et
2	Marine Le Pen	Les taxes représentent plus de 60% du prix de l'énergie. Il y a quelque chose d'inique à voir l'État augmenter ses bénéfices quand l'explosion des tarifs pèse lourdement sur le quotidien des Français. #Cavignac #MLaFrance https://t.co/RTusKIIGPZ	tax représentent plus prix énergie a quelque chose " inique voir état augmenter bénéfices quand explosion tarifs pèse lourdement quotidien français	les taxes représentent plus de du prix de l énergie il y a quelque chose d'inique à voir l état augmenter ses bénéfices quand l explosion des tarifs pèse lourdement sur le quotidien des français
3	Marine Le Pen	À la rencontre des commerçants et des Français au marché de #Cavignac. n n Les inquiétudes liées au pouvoir d'achat sont nombreuses et légitimes. n n La nouvelle hausse annoncée au 1er octobre des tarifs du gaz doit alerter tous les responsables politiques. #MLaFrance https://t.co/DEBLdaQLeF	rencontre commerçants français marché inquiétudes liées pouvoir " achat nombreuses légitimes nouvelle hausse annoncée er octobre tarifs gaz doit alerter tous responsables politiques	à la rencontre des commerçants et des français au marché de les inquiétudes liées au pouvoir d'achat sont nombreuses et légitimes la nouvelle hausse annoncée au er octobre des tarifs du gaz doit alerter tous les responsables politiques
4	Marine Le Pen	🗨️ "On doit engager un bras de fer avec les pays qui ne veulent pas reprendre leurs nationaux qui sont clandestins chez nous. On peut limiter les envois d'argent vers le pays d'origine : ces sommes sont supérieures à l'aide au développement !" @tv7_sudouest #MLaFrance https://t.co/UR2Gw0Tvz5	doit engager bra fer pay veulent reprendre leurs nationaux clandestins chez peut limiter envoi argent vers pay origine supérieures aide développement	on doit engager un bras de fer avec les pays qui ne veulent pas reprendre leurs nationaux qui sont clandestins chez nous on peut limiter les envois d argent vers le pays d origine ces sommes sont supérieures à l aide au développement

Topic Modeling

WHY TOPIC MODELING?

- To find the main topics of each campaign
- To summarize the most informative parts of communications

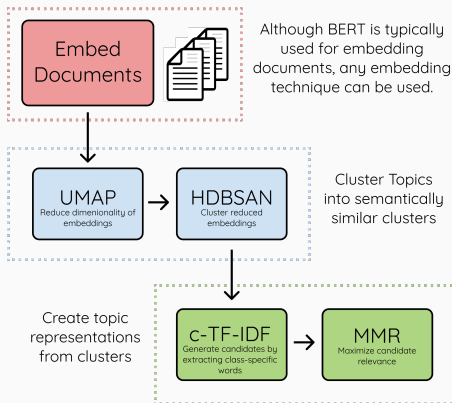
LDA

- The discretization of topics for very large or unfamiliar datasets
- bag-of-words (BOW) representation that ignore word semantics
- pre-processing is hard and does not work always

BERTopic

- BERT embeddings
- UMAP dimensionality reduction
- HDBSCAN clustering
- class based TF-IDF
- topic merging

BERTopic



DYNAMIC BERTOPIC

```

5 # Prepare data
6 lepen = pd.read_csv('clean_lepen_tweets2.csv', encoding='utf-8')
7 lepen.dropna(inplace=True)

In [2]: 1 # preparing tweet texts
        2 lepen.basic_cleaning = lepen.basic_cleaning.astype('string')
        3 lepen.basic_cleaning = lepen.apply(lambda row: re.sub(r"http[s?]", "", row.basic_cleaning).lower(), 1)
        4 lepen.basic_cleaning = lepen.apply(lambda row: " ".join(filter(lambda x: x[0] != "@", row.basic_cleaning.split())), 1)
        5 lepen.basic_cleaning = lepen.apply(lambda row: " ".join(re.sub("[a-zA-Z]+", " ", row.basic_cleaning).split()), 1)

In [3]: 1 # filtering date
        2 lepen.time = pd.to_datetime(lepen['time'], dayfirst=True)
        3 lepen = lepen[lepen['time'] > '2019-07-01']

In [4]: 1 # saving text and time in 2 different lists
        2 timestamps = lepen.time.to_list()
        3 tweets = lepen.basic_cleaning.to_list()

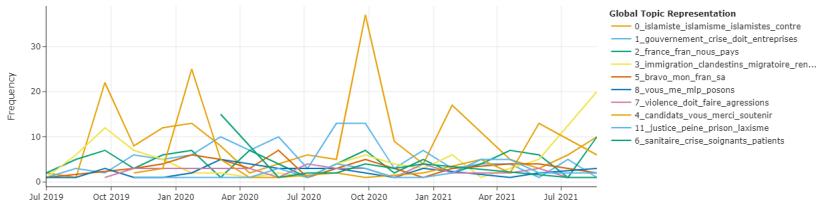
In [5]: 1 # getting topics
        2 topic_model = BERTopic(language="multilingual")
        3 topics, probs = topic_model.fit_transform(tweets)

In [6]: 1 # finding topic's distribution over time
        2 topics_over_time = topic_model.topics_over_time(tweets, topics, timestamps, nr_bins=20)
        3 # visualization
        4 topic_model.visualize_topics_over_time(topics_over_time, top_n_topics=10)

```

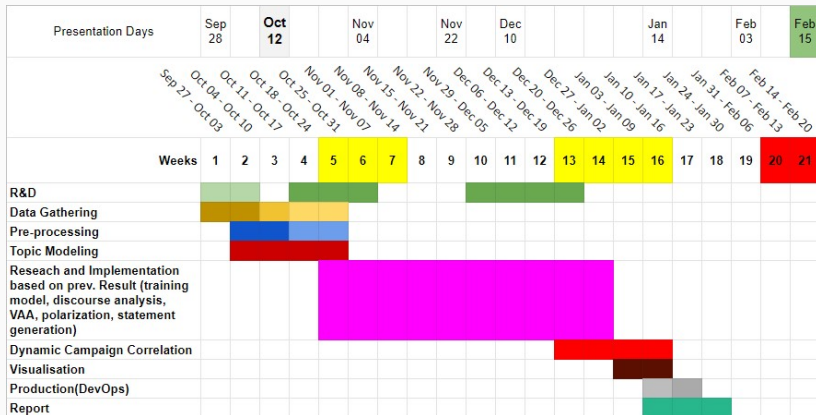
Activate Wtr
Go to Settings

Topics over Time



Project Milestones and Schedule

TIMELINE



Future Possible Approaches

FUTURE POSSIBLE APPROACHES

- ◇ Dynamic candidate affiliation
- ◇ Candidate closeness (considering VAA approaches)
- ◇ Complete statement generation

GITHUB REPO

The screenshot shows the GitHub interface for the repository 'amirpaia / election-campaign-dynamics'. The repository is public. The navigation bar includes links for Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. The main content area shows the file structure of the repository. The current branch is 'main'. There is 1 branch and 0 tags. The file list includes:

File Name	Description	Last Commit
amirpaia second presentation file		64adc30 21 minutes ago 27 commits
Data Gathering	Scripts for data collection.	3 days ago
Topic Modeling	Create bertopic.ipynb	3 days ago
articles	the standard structure of the project	15 hours ago
preprocessing	Update preprocessing_tweets.ipynb	3 days ago
presentations	second presentation file	21 minutes ago
LICENSE	Initial commit	14 days ago
README.md	Update README.md	14 days ago

Thank you!