Individual study – 6193 Spring 2018

Improvement of Twitter veridicality classifier

Contents

- 1. Overview
- 2. Phase 1
- 3. Phase 2
- 4. Phase 3
- 5. Summary

Submitted by –
Shivani Sabhlok
500237896
shivanisabhlok.1@osu.edu

1. Overview

This work is in continuation to the previously published work by Sandesh Swamy, Prof Ritter and Prof Catherine de Marneffe.

In this course, I worked closely with Prof Catherine de Marneffe. We can broadly divide the work carried out in this semester into three phases. In the following sections I will be discussing each of the phases as what we wanted to do and the analysis from each phase.

2. Phase 1

We initially started by analysis for deleted tweets (tweets that were present when we collected data before the event but were missing from the data that we collected after the event) from the Eurovision and French election datasets. We wanted to see how many tweets get deleted and what was the prediction of the classifier for those tweets.

The following were the data set features -

For Eurovision – there were 24 contenders in our dataset and in all we had 455 tweets that were collected before the event and 654 tweets that were collected after the event.

For French election – there were 2 contenders in our dataset and in all we had 1563 tweets that were collected before the event and 1672 tweets that were collected after the event.

Summary of analysis -

For Eurovision -

- On programmatically checking for tweets from the before dataset in the after-event data set, it
 was found 57 tweets were missing. To validate the results, I went back and did a hand search
 using the twitter advanced search and found that out of 57 missing tweets that were observed
 missing programmatically, 42 of them were missing from our data set that we collected after the
 event. The rest 15 were not searchable/ not found on the twitter.
- 3 of them were protected so I couldn't see their tweets. I am not sure if the tweets are available
- 2 of the accounts were suspended

For French election -

LePen

- On programmatically checking for tweets from the before dataset in the after-event data set, it
 was found 35 tweets were missing. To validate the results, I went back and did a hand search
 using the twitter advanced search and found that out of 35 missing tweets that were observed
 missing programmatically, 15 of them were missing from our data set that we collected after the
 event. The rest 20 were not searchable/ not found on the twitter.
- 6 Were from the suspended accounts
- 0 Were from protected accounts
- 2 accounts were not searchable

Macron

On programmatically checking for tweets from the before dataset in the after-event data set, it
was found 158 tweets were missing. To validate the results, I went back and did a hand search
using the twitter advanced search and found that out of 158 missing tweets that were observed

missing programmatically, 92 of them were missing from our data set that we collected after the event. The rest 66 were not searchable/ not found on the twitter.

- 25 accounts were suspended
- 2 Accounts were protected
- 14 Were not searchable

Some of the search combinations tried while manually searching using twitter advanced search –

Some of the reasons my search results on web didn't match exact -

- 1. Emoticons
- 2. Truncated youtube link
- 3. http was added to some links in our results and thus I wasn't able to match the exact phrase but removing http, I was able to find it.
- 4. Also some of the accounts were suspended.

Below are the screen shots of the missing tweets. Highlighted in red are the ones that were not found even after the manual search.

Highlighted in pink are the ones where I explicitly got the message indicating that the tweet has been deleted.

Details of the Eurovision dataset -

Before				After	
Armenia	DataFrame	(1, 6)	Column names: 0, 1, 2, 3, 4, 5	Armenia	DataFran
Australia	DataFrame	(30, 6)	Column names: 0, 1, 2, 3, 4, 5	Australia	DataFram
Austria	DataFrame	(8, 6)	Column names: 0, 1, 2, 3, 4, 5	Austria	DataFran
Azerbaijan	DataFrame	(7, 6)	Column names: 0, 1, 2, 3, 4, 5	Azerbaijan	DataFran
Belarus	DataFrame	(5, 6)	Column names: 0, 1, 2, 3, 4, 5	Belarus	DataFran
Belgium	DataFrame	(20, 6)	Column names: 0, 1, 2, 3, 4, 5	Belgium	DataFran
Bulgaria	DataFrame	(62, 6)	Column names: 0, 1, 2, 3, 4, 5	Bulgaria	DataFran
Croatia	DataFrame	(14, 6)	Column names: 0, 1, 2, 3, 4, 5	Croatia	DataFra
Cyprus	DataFrame	(4, 6)	Column names: 0, 1, 2, 3, 4, 5	Cyprus	DataFra
Denmark	DataFrame	(8, 6)	Column names: 0, 1, 2, 3, 4, 5	Denmark	DataFra
Germany	DataFrame	(1, 6)	Column names: 0, 1, 2, 3, 4, 5	Germany	DataFra
Greece	DataFrame	(8, 6)	Column names: 0, 1, 2, 3, 4, 5	Greece	DataFran
Israel	DataFrame	(7, 6)	Column names: 0, 1, 2, 3, 4, 5	Israel	DataFran
Italy	DataFrame	(4, 6)	Column names: 0, 1, 2, 3, 4, 5	Italy	DataFra
Moldova	DataFrame	(21, 6)	Column names: 0, 1, 2, 3, 4, 5	Moldova	DataFran
Netherlands	DataFrame	(11, 6)	Column names: 0, 1, 2, 3, 4, 5	Netherlands	DataFra
Norway	DataFrame	(2, 6)	Column names: 0, 1, 2, 3, 4, 5	Norway	DataFran
Poland	DataFrame	(4, 6)	Column names: 0, 1, 2, 3, 4, 5	Poland	DataFra
Portugal	DataFrame	(124, 6)	Column names: 0, 1, 2, 3, 4, 5	Portugal	DataFra
Romania	DataFrame	(22, 6)	Column names: 0, 1, 2, 3, 4, 5	Romania	DataFra
Spain	DataFrame	(6, 6)	Column names: 0, 1, 2, 3, 4, 5	Spain	DataFran
Sweden	DataFrame	(20, 6)	Column names: 0, 1, 2, 3, 4, 5	Sweden	DataFram
UK	DataFrame	(60, 6)	Column names: 0, 1, 2, 3, 4, 5	UK	DataFram
Ukraine	DataFrame	(6, 6)	Column names: 0, 1, 2, 3, 4, 5	Ukraine	DataFram

After			
Armenia	DataFrame	(1, 6)	Column names: 0, 1, 2, 3, 4, 5
Australia	DataFrame	(28, 6)	Column names: 0, 1, 2, 3, 4, 5
Austria	DataFrame	(9, 6)	Column names: 0, 1, 2, 3, 4, 5
Azerbaijan	DataFrame	(6, 6)	Column names: 0, 1, 2, 3, 4, 5
Belarus	DataFrame	(3, 6)	Column names: 0, 1, 2, 3, 4, 5
Belgium	DataFrame	(20, 6)	Column names: 0, 1, 2, 3, 4, 5
Bulgaria	DataFrame	(58, 6)	Column names: 0, 1, 2, 3, 4, 5
Croatia	DataFrame	(13, 6)	Column names: 0, 1, 2, 3, 4, 5
Cyprus	DataFrame	(3, 6)	Column names: 0, 1, 2, 3, 4, 5
Denmark	DataFrame	(6, 6)	Column names: 0, 1, 2, 3, 4, 5
Germany	DataFrame	(1, 6)	Column names: 0, 1, 2, 3, 4, 5
Greece	DataFrame	(8, 6)	Column names: 0, 1, 2, 3, 4, 5
Israel	DataFrame	(7, 6)	Column names: 0, 1, 2, 3, 4, 5
Italy	DataFrame	(204, 6)	Column names: 0, 1, 2, 3, 4, 5
Moldova	DataFrame	(21, 6)	Column names: 0, 1, 2, 3, 4, 5
Netherlands	DataFrame	(14, 6)	Column names: 0, 1, 2, 3, 4, 5
Norway	DataFrame	(2, 6)	Column names: 0, 1, 2, 3, 4, 5
Poland	DataFrame	(4, 6)	Column names: 0, 1, 2, 3, 4, 5
Portugal	DataFrame	(118, 6)	Column names: 0, 1, 2, 3, 4, 5
Romania	DataFrame	(21, 6)	Column names: 0, 1, 2, 3, 4, 5
Spain	DataFrame	(5, 6)	Column names: 0, 1, 2, 3, 4, 5
Sweden	DataFrame	(20, 6)	Column names: 0, 1, 2, 3, 4, 5
UK	DataFrame	(76, 6)	Column names: 0, 1, 2, 3, 4, 5
Ukraine	DataFrame	(6, 6)	Column names: 0, 1, 2, 3, 4, 5

missing - Dictionary (14 elemen

Key	Type	Size
Australia	list	5
Azerbaijan	list	1
Belarus	list	2
Belgium	list	1
Bulgaria	list	7
Croatia	list	3
Cyprus	list	1
Denmark	list	2
Italy	list	3
Portugal	list	10
Romania	list	5
Spain	list	1
Sweden	list	2
UK	list	14

list 2

list 1

Belarus

Belgium

Index	Туре	Size	Value
0	list	2	['I'm in Europe and it seems like no one is aware Eurovision is on Wha
1	list	2	['After listening to "sound of silence " from Australia last year I'm
2	list	2	[After watchg #SemiFinall I predict #Hovig #Gravity #Cyprus 2 #win Ca
3	list	2	'I love how they hasten to assure Europe that Eurovision won't actual
4	list	2	['Could @luciejones1 win @Eurovision for the #UK? The @7thHeavenRmx re

[['Belarusian pop duo Naviband are expected to win Best Song at Eurovi ...

[['According to the amount of downloads on #iTunes, #Belgium could win ..

Bulgaria - List (7 elements)

Index	Type	Size		Value
0	list	2	['Favorites aside, if Bulgaria wins, it'd be the 1st singer of this ce	
1	list	2	['I think I really want Bulgaria to win #Eurovision', '/scanteihood/st	
2	list	2	['Who will win this year's #Eurovision? #Italy, #Bulgaria, #Portugal o	
3	list	2	['wait bulgaria is apparently a favourite to win are you joking	
4	list	2	['I swear if Bulgaria don't win tomorrow,imma be real sad.The boy mad	
5	list	2	['I can't believe Bulgaria is going to win Eurovision for the first ti	
6	list	2	['#Eurovision Bulgaria is gonna win', '/HollisMedia/status/86297721652	

Croatia - List (3 elements)

Index	Туре	Size	Value
0	list	2	['#Croatia's #MrVoice and Hungarian gypsy singer win #Eurovision final
1	list	2	['I may actually put some money on Croatia to win, simply beautiful (a
2	list	2	['Fabulous effort. Can't wait to see you WIN on Saturday - 'douze poi

Cyprus	list	1	[['After watchg #SemiFinall I predict #Hovig #Gravity #Cyprus 2 #win C
--------	------	---	--

Denmark - List (2 elements)

Index	Туре	Size	Value
0	list	2	['After watchg #Semifinall I predict #Hovig #Gravity #Cyprus 2 #win Ca
1	list	2	['Cancel #Eurovision now, we have a winner. https://www.youtube.com/wa

Italy - List (3 elements)

Index	Type	Size	Value
muck	·ypc	Size	Yalac
0	list	2	['Sorry no matter how many times I hear Italy's #Eurovision song I sti
1	list	2	['An Italy win would make a lot of punters happy. #Eurovision http://b
2	list	2	['Italy have it in the bag. I don't really see anything special in the

Portugal - List (10 elements)

Index	Туре	Size		Value
0	list	2	['My winner of #esc2017 its undoubtedly Portugal This song makes me c	
1	list	2	['I'm not sure if we'll win but Portugal will definitely end up in the	
2	list	2	['Also pls can we talk about Portugal's entry for #Eurovision', '/give	
3	list	2	['As much as I wanted Francesco to win, it really is Portugal's year #	
4	list	2	['even if Slavador will win tomorrow (that possibility exists) 1 don't	
5	list	2	['http://ift.tt/2pGOEmp Boa Noite from Ireland! Hopefully you can win	
6	list	2	['Portugal is first on the bets to win so y'all know what to do tomorr	
7	list	2	['Who will win this year's #Eurovision? #Italy, #Bulgaria, #Portugal o	
8	list	2	['EXACTLY WHY PORTUGAL CAN'T WIN DON'T LET A BALLAD WIN PLEASE #Eurov	
9	list	2	['Don't invent these horrible things to make Portugal win.Shame on you'	

Romania - List (5 elements)

Index	Туре	Size		Value
0	list	2	okay but if Romania doesn't win #Eurovision , i will be heartbroken	
1	list	2	['can romania win eurovision my phys teacher told us she"d set us more	
2	list	2	['Who will win this year's #Eurovision? #Italy, #Bulgaria, #Portugal o	
3	list	2	['If #romania cant win this year they will not ever be able to win, ev	
4	list	2	['ROMANIA 4 THE WIN!! #Romania #Eurovision', '/IvesLovah/status/8628	

Spain list 1 [['Just a reminder Eurovision tomorrow need UK to win to show every 1 ...

Sweden - List (2 elements)

Index	Туре	Size	Value
0	list	2	['After listening to "sound of silence " from Australia last year I'm
1	list	2	['Someone really wants Sweden to win again ;) #Eurovision #ESC2017',

W UK - List (14 elements)

Index	Type	Size		Value
0	list	2	['#Eurovision actually gave the UK a promising running order position! \dots	
1	list	2	['#ESC2017: The betting odds are STILL changing - one night before the	
2	list	2	['Ooh I forgot it was Eurovision tomorrow! I do hope they make the UK	
3	list	2	['UK will get mashed at #Eurovision everyone hates Britain. Even me.',	
4	list	2	['Who do you reckon will win', '/agcu4181/status/863104575955509248']	
5	list	2	['Belgium could win Eurovision ahead of Netherlands, Italy and UK acco	
6	list	2	['Lucie Jones beautiful song beautiful singer good luck Lucie if you d	
7	list	2	['All set for #EUROVISION #UK for the win #ESC2017 @luciejones1 #12p	
8	list	2	['I'm mostly prepared for Por, Ita, Swe, Uk, Bel & Bul to win (highest	1
9	list	2	['Eurovision Song Contest 2017: Lucie Jones to WIN? UK hopeful's odds	
10	list	2	['Don't think this'll win in a million years + live performance not gr	
11	list	2	['Would LOVE for Lucie to win #Eurovision to prove the haters/doubters	
12	list	2	['Eurovision Song Contest 2017: Lucie Jones to WIN? UK hopeful's odds	
13	list	2	['Fan of UK entry Lucie Jones showcases Brexit inspired outfit ahead o	

Details of French election dataset –

Before			After			
FrenchElecBefore - Diction	nary (2 eleme	nts)	FrenchElecAfter - Dictionary (2 elements)			
Key Type	Size		Key Type Size			
LePen DataFrame	(398, 6)	Column names: 0, 1, 2, 3, 4, 5	LePen DataFrame (406, 6) Column names: 0, 1, 2, 3, 4, 5			
Macron DataFrame	(1165, 6)	Column names: 0, 1, 2, 3, 4, 5	Macron DataFrame (1266, 6) Column names: 0, 1, 2, 3, 4, 5			

FrenchElecMissing - Dictionary (2 elements)

Key	Туре	Size
LePen	list	35
Macron	list	158

LePen - List (35 elements)

Index	Type	Size		VaÎu
0	list	2	['What?????', '/kindNotWeak2017/status/861003218654031872']	
1	list	2	['Marine LePen https://twitter.com/iiovesharry/status/8609906511309455	
2	list	2	['If France wants to live and see another 18 years, they'll vote for #	
3	list	2	['@JeanMarieLepen1 marcon is use dirty trick lepen will win', '/SHIAFA	
4	list	2	['With just under 10,000 votes #Lepen took the win with over 75% of th	
5	list	2	['#Macron & #LePen the polls all predict Macron will win,and I'm predi	
6	list	2	['please RT. LePen must win #FrenchElection #MarineLePenhttps://twitte	
7	list	2	["What is at stake in the French election? http://video.foxnews.com/v/	
8	list	2	['#win LePen', '/Floridian33480/status/860978852557934592']	
9	list	2	['It will take a miracle from God above.IMHO the NWO/Globalists R 2str	

10 list 2 ['Macron qui avait une oreillette pendant le grand débat, et ca ose ac . 11 list 2 ['Who do you predict to win #frenchelection2017 #Frenchelections #Fran . 12 list 2 ['Marine Le Pen likely to win the French Presidential election. #Marin . 13 list 2 ['Americans who are not Trump supporters are praying LePen ends up in 14 list 2 ['When the peoples votes even with the conruption from the EU overwhel . 15 list 2 ['love it ,love it guys , so nice when put clips,very funny, i love @N . 16 list 2 ['I don't care, because #LePen won't win the second round.', '/Cable_S .	
12 list 2 ['Marine Le Pen likely to win the French Presidential election. #Marin . 13 list 2 ['Americans who are not Trump supporters are praying LePen ends up in 14 list 2 ['When the peoples votes even with the corruption from the EU overwhel . 15 list 2 ['love it ,love it guys , so nice when put clips, very funny, i love @N .	
13 list 2 ['Americans who are not Trump supporters are praying LePen ends up in 14 list 2 ['When the peoples votes even with the corruption from the EU overwhel . 15 list 2 ['love it ,love it guys , so nice when put clips,very funny, i love @N .	
14 list 2 ['When the peoples votes even with the concuption from the EU overwhel . 15 list 2 ['love it ,love it guys , so nice when put clips,very funny, i love @N .	
15 list 2 ['love it ,love it guys , so nice when put clips,very funny, i love @N .	
16 list 2 ['I don't care, because #LePen won't win the second round.', '/Cable_S .	
17 list 2 ['A LePen win would not surprise me a bit. Or, an early scare followe.	•
18 list 2 ['This is why LePen is going to win IMO. Crooks like Obama telling the .	
19 list 2 ['Italy must get out of the EU and the fake globalization Europe is be .	
20 list 2 ['dear @lepen , we r safe here America c/u Boss Trump on the house,&	
21 list 2 ['Difference being macron win will just be profit taking lepen will be	
22 list 2 ['yes ,yes let help @lepen win,safety and flowers for her, Vive la Fr	
23 list 2 ['dear @Nigel_Farage , I love yr voice & yr fight for UK & Europe, yes	
24 list 2 ['This Sunday will be very decisive for . Who will be elected in #fram	
25 list 2 ['She will not win, her best chance has always been 2022. France will	

26	list	2	['Full Contentious #LePen #Macron Debate (starts at Minute 25)https://
27	list	2	['The leaks will not make a jot of difference macron is going to win,
28	list	2	['That worked for Hillary didn't it? So now LePen will win.', '/ValAC
29	list	2	['Bc Putin wants LePen to win so they can form the 5thRiche, & their L
30	list	2	['.@MLP_officiel This is horrible! Macron has been working w Police &
31	list	2	['Who do you want to win the French election? LePen or Macron?', '/AzD
32	list	2	['I know nothing about #MacronLeaks. I just want #LePen to win and mus
33	list	2	['Oh no. I have a very bad feeling that LePen is going to win. Go fy @
34	list	2	['I am voting #Macron I cannot let a fascist like #LePen win.', '/Fact

Macron - List (158 elements)

Index	Туре	Size		Value
0	list	2	['2017 French Presidential election: I predicted by Chinese I Ching th	
1	list	2	['2017 French Presidential election: I predicted by Chinese I Ching th	
2	list	2	['2017 French Presidential election: I predicted by Chinese I Ching th	
3	list	2	['2017 French Presidential election: I predicted by Chinese I Ching th	
4	list	2	['I will predict PEN will win. You will see a turnaround like BREXIT a	
5	list	2	['Emmanuel Macron better the win the French Presidency, actual daddy,	
6	list	2	['This is why macron can not win https://twitter.com/TEN_GOP/status/86	
7	list	2	['Macron win means France and Europe are enslaves to the Globalists	
8	list	2	['Emmanuel Macron And Marine Le Pen Set To Win First Round Of France's	
9	list	2	['Emmanuel Macron And Marine Le Pen Set To Win First Round Of France's	
10	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
11	list	2	['Macron will win and France with rot. Only the Anglos will survive th	
12	list	2	['Most French police are with Macron and like here in America as they	
13	list	2	['@EmmanuelMacron Go Macron. Hope you win so Europe can remain united	
14	list	2	['Sure she is fascist. But she won't win, your desperate efforts to he	
15	list	2	[Not to give Le Pen the win, but to have Macron starting out his term	
16	list	2	['#Macron & #LePen the polls all predict Macron will win…and I'm predi \dots	
17	list	2	['Check out how these lunatics think the free press should allow a glo	
18	list	2	('Emmanuel Macron claims massive hack as France prepares for Sunday vo	
19	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	

Index	Туре	Size		Value
20	list	2	['lmao the Macron campaign put out a video of US pundits predicting Hi	
21	list	2	['Macron qui avait une oreillette pendant le grand débat, et ca ose ac	
22	list	2	['#Presidentielle2017 bets in UK going on Le Pen. At 15/2 that has to	
23	list	2	['Morgen de Franse Verkiezingen hehe Easy win voor Macron', '/_Rick_M	
24	list	2	['Although I'm not a fan of Macron either, he MUST win to prevent Mari	
25	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
26	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
27	list	2	['Le Pen and Macron same words but different meanings both put there t	
28	list	2	['Who do you predict to win #frenchelection2017 #Frenchelections #Fran	
29	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
30	list	2	['Shoutout to Macron for helping me win 700 bucks,', '/powerpuffsama2/	
31	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
32	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
33	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
34	list	2	['Brexit was 100% the correct decision over here. Macron does NOT care	
35	list	2	['Macron is a little boy. He's knows very little and that will become	
36	list	2	['Since when is the West required to import people who want to destroy	
37	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
38	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
39	list	2	['What executive powers will Emmanuel Macron or Marine Le Pen have if	

	_			
	-71-	Size		Valu
40	list	2	['Macron was alway going to win. It's a shame for them, so the leaks w	
41	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
42	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
43	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
44	list	2	['Macron will win now because French media blacked it out and macron w	
45	list	2	['The secures a Macron win and gives more Global leverage against for	
46	list	2	['People who claim #CIA hacked #Macron, who they want to win, to damag	
47	list	2	['But the laws do not apply across the EU so we need to see this being	
48	list	2	['Macron is a bought and paid for world banker Rothchild stooge.', '/T	
49	list	2	['A LePen win would not surprise me a bit. Or, an early scare followe	
50	list	2	['They aren't falling for it, I hear. They see what we're going throug	
51	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
52	list	2	['Macron might lose the election despite polls and his early lead. Wil	
53	list	2	['Go long euro #Frenchelectionsmacron might win', '/suman_rd/status	
54	list	2	['You win popular vote you win the election.', '/MacronTV/status/86086	
55	list	2	['Who doesn't know rich people cheat on taxes. Macron will win, do not	
56	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
57	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
58	list	2	['OMG j'ai oubliéde voter pour macron.En Marche pic.twitter.com/ZZo	
59	list	2	['I throw my support under #Macron #ThingsJesusNeverSaid #MacronLea	

Index	Type	Size		Value
60	list	2	['love it love it, Macron team ,deleting evidence, guys very funny, le	
61	list	2	['Macron will win but will he last a fortnight?', '/Paulxxxxx/status/8	
62	list	2	['Difference being macron win will just be profit taking lepen will be	
63	list	2	['Yes macron win is the rumour but either result I believe will result \dots	
64	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
65	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
66	list	2	['Money is pouring in on Macron to win the French election http://news	
67	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
68	list	2	['Money is pouring in on Macron to win the French election http://dlvr	
69	list	2	['Money is pouring in on Macron to win the French election http://dlvr	
70	list	2	['Money is pouring in on Macron to win the French election http://dlvr	
71	list	2	['Tomorrow: If Le Pen wins I'll be horrified. If Macron wins I'll be	
72	list	2	['2017 French Presidential election: I predicted by Chinese I Ching th	
73	list	2	['France will fall. Macron will win and the death of France and its pe	
74	list	2	['2017 French Presidential election: I predicted by Chinese I Ching th	
75	list	2	['2017 French Presidential election: I predicted by Chinese I Ching th	
76	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
77	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
78	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
79	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	

Index	Type	Size		Value
80	list	2	['Not at all macron, like Chirac, will win because many can't vo	
81	list	2	['2017 French Presidential election: I predicted by Chinese I Ching th	
82	list	2	['The globalists will not allow Le Pen to win in France. Social media	
83	list	2	['2017 French Presidential election: I predicted by Chinese I Ching th	
84	list	2	['Dear @BuzzFeed: It wasn't France that got hacked, only Macron & Co	
85	list	2	['2017 French Presidential election: I predicted by Chinese I Ching th	
86	list	2	['AFP: What executive powers will Emmanuel Macron or Marine Le Pen hav	
87	list	2	['We have not yet discovered fakes in #MacronLeaks & we are very skept	
88	list	2	['France's tradition of liberty has never been under threat like now	
89	list	2	['This Sunday will be very decisive for . Who will be elected in #fran	
90	list	2	['Also, it appears that #MacronLeaks contain badly forged fakes. And t	
91	list	2	['But Macron is in the lead! Le Pen won't win! https://www.youtube.com	
92	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
93	list	2	[*FRANCE do NOT believe the polls that say MACRON will win. Come out t	
94	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
95	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
96	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
97	list	2	['Full Contentious #LePen #Macron Debate (starts at Minute 25) https:/	
98	list	2	['The leaks will not make a jot of difference macron is going to win,	
99	list	2	['no chance : the effect of the last fakenews this week, will be a big	

Index	Туре	Size		Value
100	list	2	['E. Macron: "After I WILL win, I will give to France THE BIGGEST Fren	
101	list	2	['Sadly fillon got killed for similar . Macron lied to the public and	
102	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
103	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
104	list	2	['The hashtag was shutdown! Macron will win no matter what!', '/NYCisp	
105	list	2	['Can any one produce an arguement on why russia would prefer le pen t	
106	list	2	['France should remember this when they vote: #Macron is a 'non-natio	
107	list	2	['As an American I would have to agree French are not as gullible as m	
108	list	2	['Have u heard about Posobiec, Cerno, 4chan breaching French law w Mac	
109	list	2	['So, basically the #MacronLeaks are real, but there's nothing interes	
110	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
111	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
112	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
113	list	2	['.@MLP_officiel This is horrible! Macron has been working w Police &	
114	list	2	['Who do you want to win the French election? LePen or Macron?', '/AzD	
115	list	2	['Libertarians should definitely want Macron to win. In French terms h	
116	list	2	['Russia didn't get to Macron in time because they didn't know if he'd	
117	list	2	['Great distraction campaign by focusing on who hacked #macron rather	
118	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
119	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	

Index	Туре	Size		Value
120	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
121	list	2	['Money is pouring in on Macron to win the French election https://goo	
122	list	2	['Kind of wanted Le Pen to beat Macron fair in square but a win's a wi	
123	list	2	['Macron having an each way bet.Frighten voters with russian menace, b	
124	list	2	['As bitter French campaign ends, Macron's team hit by hack http://a.m	
125	list	2	['New post (Money is pouring in on Macron to win the French election)	
126	list	2	['businessinsider: Money is pouring in on Macron to win the French ele	
127	list	2	['I can't find any of the Macron Emails online. Does anyone know what \dots	
128	list	2	['New post (Money is pouring in on Macron to win the French election	
129	list	2	['Money is pouring in on Macron to win the French election Read more:	
130	list	2	['RT: businessinsider:Money is pouring in on Macron to win the French	
131	list	2	['Macron sucks shit and everyone is patting themselves on the back for \dots	
131 132	list list		['Macron sucks shit and everyone is patting themselves on the back for ['Despicable. A must win game and throw it away. That joke that came i	
		2		
132	list	2	['Despicable. A must win game and throw it away. That joke that came i	
132 133	list list	2 2 2	['Despicable. A must win game and throw it away. That joke that came i ['En même temps elle fera intermédiaire pendant que Macron lui sera lu	
132 133 134	list list list	2 2 2 2	['Despicable. A must win game and throw it away. That joke that came i ['En même temps elle fera intermédiaire pendant que Macron lui sera lu ['Macron Campaign Says It Was Target of 'Massive' Hacking Attack https	
132 133 134 135	list list list list	2 2 2 2 2	['Despicable. A must win game and throw it away. That joke that came i ['En même temps elle fera intermédiaire pendant que Macron lui sera lu ['Macron Campaign Says It Was Target of 'Massive' Hacking Attack https ['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w	
132 133 134 135 136	list list list list	2 2 2 2 2 2	['Despicable. A must win game and throw it away. That joke that came i ['En même temps elle fera intermédiaire pendant que Macron lui sera lu ['Macron Campaign Says It Was Target of 'Massive' Hacking Attack https ['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w ['Russian bots and stupid MAGAs are at work once again. But they won't	
132 133 134 135 136 137	list list list list list	2 2 2 2 2 2 2 2	['Despicable. A must win game and throw it away. That joke that came i ['En même temps elle fera intermédiaire pendant que Macron lui sera lu ['Macron Campaign Says It Was Target of 'Massive' Hacking Attack https ['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w ['Russian bots and stupid MAGAs are at work once again. But they won't ['And now Macron is trying the same ploy', '/platozkv/status/860667411	

140	list	2	['Emmanuel Macron adviser urges French to vote amid 'danger' Marine Le
141	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w
142	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w
143	list	2	['The globalists are already making excuses for why Macron is going to
144	list	2	['macron is a pin head LE PEN for the WIN! https://twitter.com/cnnbrk
145	list	2	['It's not fake. Wikileaks confirmed its authenticity and have found n
146	list	2	['I am voting #Macron I cannot let a fascist like #LePen win.', '/Fact
147	list	2	['Since when has #neolibfail been "centrist". If you want more #global
148	list	2	['Le Pen is so bad that she can't win even though French voters know M
149	list	2	['Emmanuel Macron And Marine Le Pen Set To Win First Round Of France's
150	list	2	['This is what will happen in France and why! Macron win pic.twitter
151	list	2	('Or maybe the allegations are just attempts by Macron to Win Votes',
152	list	2	['#MacronLeaks #Macron is now under a serious investigation. The Ca
153	list	2	['Next up - Le Pen for the shock winMacron to blame Russia for ste
154	list	2	['Macron win not help euro for now', '/zeekhan0926/status/860646282964
155	list	2	['I want Macron to WIN so that French people will start seeking asylum
156	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w
157	list	2	['LADBROKES-FRENCH PRESIDENTIAL ELECTION 6/1 for Emmanuel Macron to w

3. Phase 2

Determining the missing tweets, led us to the phase 2 of the project where we were interested in knowing if the deleted tweets were positive/negative or neutral.

I ran the classifier on the missing tweets listed above and also manually annotated the tweets as positive/neutral/negative.

Most of the tweets were found to be neutral. So, we realized that tweet deletion was not a significant issue. But while manually annotating the tweets, it was observed that some of the tweets which were clearly negative/positive were incorrectly predicted to be neutral and vice-versa.

This led us to the third phase of the project where we started analyzing why the seemingly obvious classes of the tweets were misclassified.

4. Phase 3

In this phase, we were trying to analyze why some of the clearly positive/negative tweets were incorrectly predicted neutral and vice-versa.

For doing this analysis, I picked a set of interesting tweets, as I call it, and analyzed the generated features for these tweets.

Below is a screen shot of few of the interesting tweets –

tweet	contender	predicted	gold
hat worked for Hillary didn't it? So now LePen will win.		neutral	positive
I will predict PEN will win. You will see a turnaround like BREXIT and TRUMP win. And if MACRON WINS don't u go to THE 751 NO GO ZONES.	Macron	neutral	positive
This is why macron can not win https://twitter.com/TEN_GOP/status/860992055706882051	Macron	neutral	negative
Macron win means France and Europe are enslaves to the Globalists. The last stand for Western civilization will be in America		positive	neutral
@EmmanuelMacron Go Macron. Hope you win so Europe can remain united.	Macron	positive	neutral
Shoutout to Macron for helping me win 700 bucks.	Macron	positive	neutral
Since when is the West required to import people who want to destroy it? Macron is an idiot who (unfortunately) will win a rigged election	Macron	neutral	positive
Difference being macron win will just be profit taking lepen will be a dent size dent	Macron	positive	neutral
Also, it appears that #MacronLeaks contain badly forged fakes. And there's a flood of Twitter/FB bots spewing anti-Macron comments.	Macron	positive	neutral
France should remember this when they vote: #Macron is a 'non-nationalist' = Globalist Le Pen will win pic.twitter.com/UGwQcsRvkl	Macron	positive	negative
Russia didn't get to Macron in time because they didn't know if he'd win and only had a two week window to sink him.	Macron	positive	neutral
Money is pouring in on Macron to win the French election Read more: http://firenewsfeed.com/finance/35026 #money #bettingOdds #MarineLePenpic.twitter.com/EXM9RCvUlh	Macron	positive	neutral
The globalists are already making excuses for why Macron is going to lose. Hillary 2.0, they said she would win in a landslide. Nope.	Macron	positive	neutral
macron is a pin head LE PEN for the WIN! https://twitter.com/cnnbrk/status/860655554385108996	Macron	positive	neutral
#MacronLeaks #Macron is now under a serious investigation. The Cayman Island Coconut business story hasnt worked #Presidentielle2017	Macron	positive	neutral
Next up - Le Pen for the shock winMacron to blame Russia for stealing the electionrinse and repeat.	Macron	positive	negative

I analyzed generated features, the features learned by the classifier and the weights for the features.

On a high level, following is the classifier design –

It processes each tweet and starts with tokenizing the tweet. The URLs and the quotes are removed from the tweet. The classifier then works to generate features of the tweet like presence of keyword 'win' in the tweet, the distance of the keyword from the TARGET/OPPONENT, if the keyword has been negated etc. This list of features is passed to a method that reads through the trained features dictionary. Based on which of the generated features have been previously seen by the classifier (that is, exist in the learned features), a Boolean array flags the matched features. The LR multiclass classifier adds up weights of the features and multiplies with their respective co-efficient and finally predicts the probability of each class and outputs the final class as the one which has the highest probability.

I worked with the code in the demo folder on Github. The 'demo.py' has the code to stream tweets and process them whereas 'runclassifier.py' works just fine if you have already gathered the tweets and are interested in just classifying them. 'runclassifier.py' takes 2 arguments – the event file and the tweet file. Even though the details of these are present in the ReadMe.md on Github, I will still summarize here –

In the 'event.txt' (given as first argument to 'runclassifier.py'), the first line is the name of the event, the second line is the name of the TARGET and the rest of the lines have names of OPPONENTs. Make sure that the 2nd argument (file containing tweets) has the tweets that were gathered by querying 'win' for the TARGET (the contender in the second line of the event file).

The tweet file is expected to be in the following format –

Tweet, Timestamp, Username, URL

To run the classifier on the previously collected tweets, you can run the following command –

python runclassifier.py event.txt tweets.csv > output.txt

Following is the summary of the observations and the discussions around improvement of the classifier-

Note – I worked only with the trained features file that was accessible on Github at that time. So all the observations are based on that.

- 1. The learned features were in mixed case where as we work with only lower-case tweet. So many features, even though they were learned, didn't match when checked in the trained features dictionary. This might be because during the entire processing, tweet is converted to lower case in methods. Ideally, the tweet should be changed to lower case immediately after reading it. No case conversions should happen inside helper methods.
- The spaces around the keyword in the learned features was inconsistent. So even if the case matched, the spaces didn't match. So, lot of features were missed while actually classifying the tweet.
- 3. Some of the helper methods like the ones that check the distance of keyword from the target and the method that checks for dangling negation have a check

"if keyword in tweetWord"

This condition is true even if tweetWord is" swing", "knowing", "growing" etc.

This leads to incorrect feature generation.

4. Other helper methods have a check

"if keyword == tweetWord"

5. Observing the above two conditions in the code, I realized there should be a consistent kind of check for the keyword for generating features. Also, to improve the classifier, the potential variations of the keyword like "winner"," wins"," winning" etc. should also be considered as keywords.

On going back to the learned features, following were observed to be suitable variations of the keyword –

must-win

win

wins

winner

winning

winnin

winners

#win

To test this, I went back to the datasets mentioned in the first section plus an additional dataset for the Oscars 2018. On processing the total of around 7000 tweets, using the above-mentioned set of words as keywords, I could process 62 tweets more and didn't miss any good tweet.

We also analyzed the parse tree features that were generated. Take 2 sample tagged tweets –

"TARGET1 will win"

"TARGET1 will not win"

For both the tweets, the tweeboParser would make "will" to be the root of the parse tree. The helper method looks for the shortest path between the keyword and the target and thus for both the above tweets, the generated parse tree feature will be

'TARGET1 <- will -> KEYWORD'

Even though for the second tweet, the other helper method will set the flag for presence of negation but we discussed that adding "NEG" on the parse tree for the second tweet would help the classifier confidently predict the tweet negative.

- 7. It might not be feasible, but it was discussed as a nice to have feature. While working with the French election dataset, I observed that lot of times "LePen" was referred as "Le-Pen" or "Le Pen". Thus we missed out on tweets that used anything other than "LePen". So we discussed that in the events file, we could give an option to give csv for alternate name references of a contender.
- 8. Since we are dealing with user typed text, it was discussed that normalizing text might help.
- 9. Looking at learned features, it was observed that "@someuser" was also learned. This is not a helpful feature to learn. This can easily be pruned because this is tagged as "USR".

5. Summary

We discussed that if the training and the test data used for the paper along with the code is made available, the observations and analysis summarized in phase 3 could be verified and confidently worked upon.

Misc -

I think it's a bug. While processing the tweets I had, the 'getsegments' method in 'runclassifier.py' threw exception and I changed the line of code. This line would exist in 'demo.py' as well. 'runclassifier.py' was modified with 'demo.py' as the base version. I modified original line(294/298) to line(292/296)

```
def getsegments(tweet, annots, tag, lower=False, getIndices=True):
274
          result = []
275
          start = None
276
          for i in range(len(tweet)):
277
               if annots[i] == "B-%s" % tag:
278
                  if start != None:
     279
                      if getIndices:
280
                          result.append((' '.join(tweet[start:i]), (start,i)))
281
     else:
282
                          result.append(' '.join(tweet[start:i]))
283
                  start = i
284
     自
              elif annots[i] == '0' and start != None:
     285
                  if getIndices:
286
                      result.append((' '.join(tweet[start:i]), (start, i)))
     287
288
                      result.append(' '.join(tweet[start:i]))
289
                  start = None
290
     中
          if start != None:
291
     if getIndices:
292
                  result.append((' '.join(tweet[start:len(tweet)]), (start, len(tweet))))
293
                  ##changed by shivani
294
                  #result.append((' '.join(tweet[start:i]), (start, i)))
     295
              else:
                  result.append(' '.join(tweet[start:len(tweet)]))
296
297
                  #changed by shivani
298
                  #result.append(' '.join(tweet[start:i]))
299
         if lower:
300
              if getIndices:
301
                  result = [(x[0].lower(), x[1]) for x in result]
302
303
                  result = [(x.lower()) for x in result]
304
         return result
```

I would like to thank Prof Marie for taking me as her student for this course and giving me helpful insights every week. I have definitely learned a lot from this project and her continuous guidance.

I would also like to thank Sandesh Swamy for helping me understand this project better and sharing the data for the analysis.