Big Data Analytics for Opinion Mining and Patterns Detection of the Tunisian Election

Zeineb Dhouioui
Bestmod Laboratory
ISG Tunis
University of Tunis
zeineb.dhouioui@hotmail.fr

Hanen Bouali
Bestmod Laboratory
ISG Tunis
University of Tunis
hanene.bouali@gmail.com

Jalel Akaichi
Bestmod Laboratory
ISG Tunis
University of Tunis
j.akaichi@gmail.com

Abstract

Big Data refers to an enormous volume of structured and unstructured data that cannot be handled with traditional databases. The emergence of big data is due to the huge quantities of information. Currently, researchers tend to study big data particularly data produced from social networks. Nowadays, these latest become a vital and crucial tool in tracking and extracting public opinion for developed countries offering benefits to the democratic process of election. In this paper, we aim to identify the political preferences and tendency of the Tunisian population using classification and opinion mining techniques. To prove the usefulness of the proposed method, we analyze electoral data sets in Tunisia obtained from the official sites of independent higher instance for election. This analysis demonstrates close correspondence between election results and extracted opinion.

1 Introduction

Big data is an extensively used term in tremendous data miscellany. Thus, this huge amount of data makes hard and sometimes impossible the analysis in a comfortable way using traditional data processing techniques (Anjaria and Guddeti, 2014). The dares includes pattern recognition, analysis, prediction ... Given the benefits of big data, they afford a chance to understand collected data in order to predict data patterns. With millions of people using social networks to express their opinions, a tremendous volume of data is generated. Unfortunately, elections data that was published by the ISIE present only a summary of the vote counts. Detailed data is not accessible. Combining social networks data and ISIE data, voter behavior can be defined. (Sudhahar et al., 2015)

With the evolution of the web 2.0 and the large number of users, big data analysis becomes very useful in pattern detection. Particularly, we handle in this paper Tunisian Voters in legislative and presidential election. Indeed, despite the mutual relation between socio-economic characteristics and voters in Tunisian elections, for the best of our knowledge, no researchers treat this task. Moreover, Elections in countries on path to democracy are considered as a source of enormous quantity of data.

The exponential rise of social networks popularity influences real-world politics. These platforms have been exploited in Arab revolution as a mobilization tool and showing social movements. Moreover, social networks are used to study voters preferences in order to forecast electoral results. We cant ignore the role of social networks especially Facebook in our election, using opinion mining we can analyze and predict the voting intention of Facebook users in Tunisia for the 2014 presidential election. The huge amounts of public opinion data need accurate tools to extract useful information.

In this study, we exploit stables and some statistical official data from ISIE in order to identify possible correlations between socio-economic features of voters and elections results. Furthermore, to well draw voting behavior, we study psychological sociological and economic variables. Thus we have discovered high correlation. The exploited data consists of the results of the Tunisian election of 2014 and 2011, also as we said previously some socio-economic variables and some data from social networks especially Facebook and Twitter. In fact, we treat Facebook activity data during months preceded the legislative and presidential elections. An important part of our study is to determine positive and negative opinions. These opinions can then be interpreted to define the political position of voters.

Persons who share several common characteristics present a tendency to have common voting behavior. We are interested in establishing voting behavior according to voter′s profile.

The remainder of this paper is as follows: in part 2, we discuss existing works handling Socio-economic variables affecting elections results and voter′s sentiment analysis. In part 3, we present data and methodologies use in this paper in order to present opinion mining patterns of voters in the Tunisian election in part 4. Finally, an overview of our work and opportunities that has been opened up are provided in part 5.

2 Related Works

Before presenting the use case of this paper, we flew over existing works handling opinion mining voters patterns. Two sides were manages:

- Patterns affected by socio-economic variables
- Voter′s sentiment analysis

2.1 Socio-Economic Variables

The paper develops in (Kreuzer and Pettai, 2003) an analytical framework, which incorporate politician driven inter-party mobility and voter induced electoral changes. The proposed framework was applied in Estonia, Latvia and Lithuania. Authors develop different strategies and patterns which are:

- Organizational Affiliation Strategies:
  - Staying Put
  - Party Switching
  - Fusion
  - Fission
  - Starting Up
- Electoral Affiliation Strategies:
  - The frequency of voters affiliation changes
  - Voters preferences among for different party origin
- Patterns of Party Systems Transformation:
  - Alignment patterns
  - Re-alignment patterns
  - De-alignment patterns
- Organizational Affiliation Patterns of Politicians

- Electoral Affiliation Patterns of Voters:
  - Parties having won at least one seat
  - Parties emerging from break away legislative
  - Parties whose successor or predecessor (through fusion or fission) won at least one seat

Other patterns were detected in (Akarca and Tansel, 2007) which study the voters′ behavior in Turkey. Authors found that Turkish voters take economic performances into account and voters exhibit a tendency to vote against parties holding power. In addition to that, party preferences of Turkish depend on their socio-economic characteristics. Moreover, Voters distinguish between major and minor parties and hold party accountable for economic growth. The growth rate more than a year before an election does not affect its outcome. Later, a comparative analysis of voter turnout in regional elections is made in (Henderson and McEwen, 2010). This paper conducts a cross sectional examination of voter participation in regional elections in nine states between 2003 and 2006. Authors found that variations in the strength of political autonomy and the strength of attachment to the region among the electorate have a strong and positive impact on the level of turnout in regional election. The hypothesis of voter turnout in regional election is higher in regions with a high degree of regional distinctiveness was proved. Other works were done to see the impact of weather on the voter turnout in (Persson et al., 2014) .The case of Swedish election system was studied, authors cannot find a robust and statistically significant negative effect of Election Day rain turnout, unlike in the US one inch of rain reduces turnout with about one percent.

2.2 Sentiment Analysis

We should firstly define sentiment analysis called also opinion mining, it is considered as a new text analysis (Choy et al., 2011). In fact, sentiment analysis is a technique serving to analyze people′s opinions, sentiments, appraisals, and emotions. There have been several works in this area, in (Choy et al., 2011) authors discussed how sentiment analysis can predict presidential election results in Singapore according to the conversion of Twitter data. In (Ceron et al., 2015), the presented methodology consists on applying supervised method to analyze the voting intention of Twitter users in the United States during the 2012 presidential election and for the two round of the centre left2012 primaries in Italy. Authors (Jungherr et al., 2012) check out whether Twitter affects the information exchange about political affair. They also present an evaluation about Twitter messages which may reflect political sentiment. Moreover, they analyze Twitter activity in order to predict parties popularity. In (Conover et al., 2011), the main idea was to apply latent semantic analysis to tweets, and thus they discover hidden structure. Later, Lei et al in [10] propose a model able to predict the public opinions about republican presidential elections. Finally, in (Ceron et al., 2014) authors assert that sentiment analysis gives accurate predictions and can become useful in polls. Moreover,authors in(Mohammad et al., 2014) automatically annotate a set of 2012 US presidential election tweets for a number of attributes pertaining to sentiment, emotion, purpose, and style by crowd sourcing. They show through an analysis how public sentiment is shaped, tracking public sentiment and polarization with respect to candidates and issues, understanding the impact of tweets from various entities. Other research such as(Dridi, 2015) present a system for finding, from Twitter data, events that raised the interest of users within a given time period and the important dates for each event. In order to select the set of events, they used three main criteria: frequency, variation and Tf-Idf.

3 Data and Methodology

To reach our objective, we have chosen the results of the 2014 parliamentary and presidential elections in Tunisia. The election results were fairly contested and were not surrounded by any extraordinary events. The sources of our data are the web site of the Independent Higher Instance for Election (ISIE) 2014 and the National Institute of Statistic (INS). The two references report the data from 24 governors and 264 delegations for the legislative, the first and second round in presidential election. We collected also Twitter and Facebook information from the start of the legislative election on October 2011 and the electoral campaign from August 2014 to the end of second round of presidential election. Political tweets and Facebook comments includes those coming from followers and candidates. To analyze data collected from ISIE and INS, we apply WEKA algorithms. A study in (Bouali and Akaichi, 2014) has been done and proves the efficiency and efficacy of Support Vector Machine. The first step is to load an excel file to Weka. Second, we select SVM as technique. Then, in the training set, we choose the method of cross validation and we select classification′s attribute. Later on, to extract the sentiment of the collected data, we use opinionFinder. We proceed first for the data preprocessing. Second, we identify sentiment expression in the text document. Then, we detect two sets of expressions: those written in formal languages (English, Arabic and French) and those in informal languages. Thus, the first set is handled using SentiStrength tool and the second one manually. Emoticons significance is also detected using SentiStrength. In this section we propose a general method; this framework is divided into 3 steps:

- Step 1 : Data Collection : status are collected according to the most frequently used keywords. Status containing insults are removed, and status written in different languages are handled manually.
  People discus and express their opinion by commenting, posting or sharing status (or tweets). We are interested only to status, tweets and comments containing keywords related to election, parties and candidates manually found. People publications may contains emoticons and informal language, thus we should analyze those manually. After finding appropriate status or comments referring to our subject, the collected data consists of 274 status and 586 comments.

- Step2 : Status Classification
  Counting the number of positive, negative and neutral messages. Actually, using opinionFinder tool we are able to extract automatically subjective sentences and sentiment expressions. Based on SentiStrength which estimates the strength of positive and negative sentiment in texts and emoticons in various languages, we classify words into categories.
  To design the voter profile, our system is based on identifying his sentiment. Consequently, we can conclude if a given voter is with or against party or candidate. Additionally, we can interpret his preferences according to his opinion.

The aim of this work is to determine whether social networks data, and the data gathered from the ISIE official site can predict election results. The figure below illustrates the proposed methodology 1:

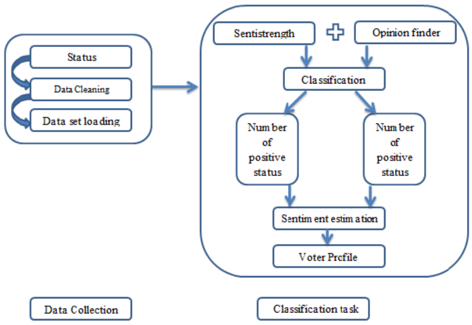

Figure 1: The System Architecture

4 Empirical Results

4.1 Voter Behavior

The objective of this paper is to extract relevant and interesting knowledge from big amount of data using supervised learning techniques. when applying our chosen methodology, our important finding is that voters distinguish between major and minor parties in a governing coalition and hold only the major incumbent party accountable for economic growth.

There is a tendency for parties to lose or win portion of their votes between first and second round due to political mergers which eventually happen in the second round. This refers to voter turnout patterns which are:

- Staying put: a voter remains affiliated with his her party in legislative and both rounds of presidential election.
- Party Switching: different from fusion (party's voters re-affiliates in the form of a merger with another party).

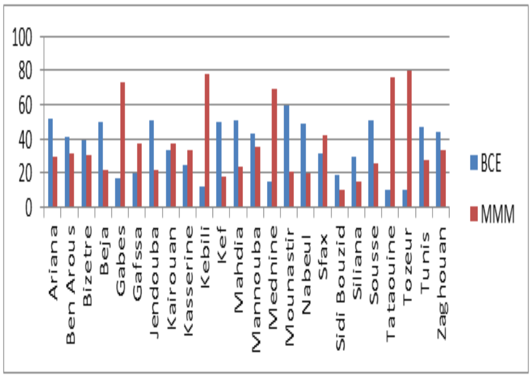

Figure 2: Votes by Governorate

- Starting up: unaffiliated voters change their opinion to be affiliated to another party. We have found that Islamist party represented by Mohamed Moncef Marzouki (MM receives their votes from people living in areas suffering from lack of economic growth such as Sidi Bouzid, Kebili... NIDAA TOUNESS represented by Beji Caied Sebssi (BCE) received their votes from those living in regions experiencing good economic conditions and economic growth like Tunis, Nabeul 2.

The recent election shows a decrease in NAHDHA (Political Party) partisans instead of an increase in NIDAA TOUNESS(Political Party). This is due to several reasons:

- A tendency to vote against the parties holding power: voters who votes BCE to decrease the power to MMM.
- Government's economic performance
- Decrease of Quality of life
- Increase of poverty rates.

We observed that citizens of a minor socio-economic status don't exhibit political efficacy

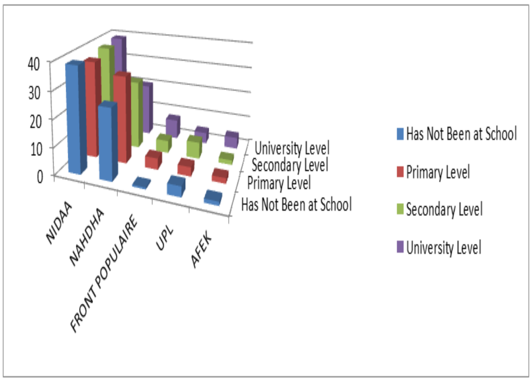

Figure 3: Education Level

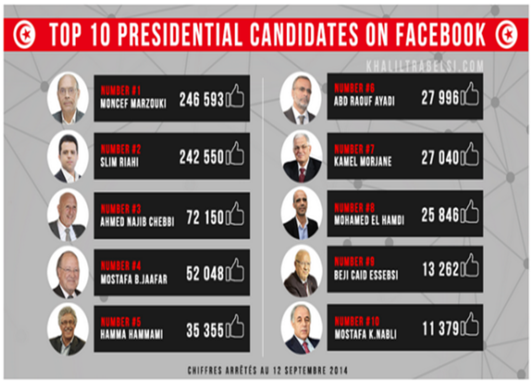

Figure 4: Number of Facebook users

and commonly ”participate less in the voting process”. In contrast, citizens with important socio-economic status participate and sponsor the political process. Thus, voter′s income and occupation are crucial in vote′s process. We mention that high education level and low level of illiteracy are behind significant civic activity 3. We found also that woman voted more to BCE.

4.2 Voter Sentiment

Tunisian presidential election 2014 in the second round was limited between two main candidates where the concurrence was severe and intense. Actually, most sentiment and opinion have been predicted and discovered according to online discussions, status and comments. Collecting and analyzing sentiment is essential for electoral campaigns. Indeed, the enormous quantity of data related to public sentiment and opinion is considered as big data which played a crucial role for Obama campaign.
With the accessibility of social networking sites, internet users can easily express their opinions about various topics. Thus, we can exploit and mine the shared opinions. In this work, we collect data from Facebook, we focus on the content of status containing a reference to political party or also politician(Mehndiratta et al., 2014).

In the following table, we present an example of positive, negative and neutral tweets. The figures below (figure 4 and 5) provide a curve that mentions the evolution of the number of tweets between August 15 and September 14, 2014; we recorded 2,004 tweets containing the hashtag TnElec. This hashtag is sometimes used more than 150 times a day for the last few weeks and also an overview about the number of Facebook users that like the candidates official pages. The x axis represent the number of negative and positive statues written for both presidential candidates The number of negative status for MMM is larger comparing to BCE according to the collected data (figure 6) which reflects the reality i.e.

Opinion Nature	**Tweet Example**
Positive	Financial Times: L’lection de Caid Sebssi est une chance pour la Tunisie
Negative	Des tunisiens craignent un retour en arrire aprs lection de BCE!
Neutral	BCE: Rencontre avec les prsident des clubs sportifs

Table 1: Example of Tweet Classification.

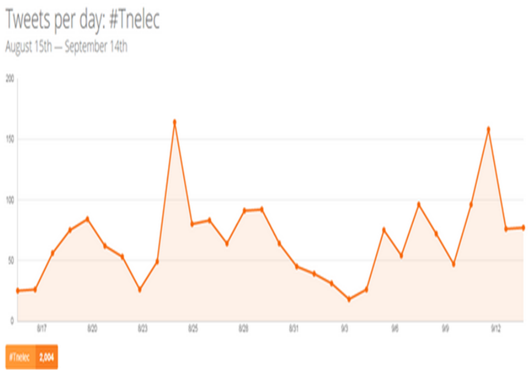

Figure 5: Tweets number evolution

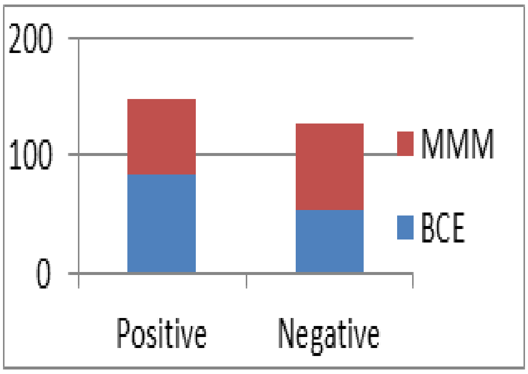

Figure 6: Comparative sentiment variation for BCE, MMM (Status)

the elections results.

5 Conclusion

Social networks have become a potential source and a popular practice during election process. We used sentiment analysis techniques to detect correlations and voters behaviors. Our statistical analysis of the 2014 Tunisian election results and the social, economic and political conditions surrounding it, leads us to conclude that: our results show that the party choice of Tunisian voters is affected by economic, demographic and socio-economic characteristics. We are aware that some interpretations require additional data. Interpretations presented in this work can serve as hypotheses in next elections. In bulk, despite the limits of social networks analysis, we can assert that our results are encouraging.

References

Ali T Akarca and Aysit Tansel. 2007. Social and economic determinants of turkish voter choice in the 1995 parliamentary election. *Electoral Studies*, 26(3):633–647.

Malhar Anjaria and Ram Mohana Reddy Guddeti. 2014. Influence factor based opinion mining of twitter data using supervised learning. In *2014 Sixth International Conference on Communication Systems and Networks (COMSNETS)*.

Hanen Bouali and Jalel Akaichi. 2014. Comparative study of different classification techniques: Heart disease use case. In *Machine Learning and Applications (ICMLA), 2014 13th International Conference on*, pages 482–486. IEEE.

Andrea Ceron, Luigi Curini, Stefano M Iacus, and Giuseppe Porro. 2014. Every tweet counts? how sentiment analysis of social media can improve our knowledge of citizens political preferences with an application to italy and france. *New Media & Society*, 16(2):340–358.

Andrea Ceron, Luigi Curini, and Stefano M Iacus. 2015. Using sentiment analysis to monitor electoral campaigns method mattersevidence from the united states and italy. *Social Science Computer Review*, 33(1):3–20.

Murphy Choy, Michelle LF Cheong, Ma Nang Laik, and Koo Ping Shung. 2011. A sentiment analysis of singapore presidential election 2011 using twitter data with census correction. *arXiv preprint arXiv:1108.5520*.

Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011. Predicting the political alignment of twitter users. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 192–199. IEEE.

Houssem Eddine Dridi. 2015. Détection d'évènements à partir de twitter.

Ailsa Henderson and Nicola McEwen. 2010. A comparative analysis of voter turnout in regional elections. *Electoral Studies*, 29(3):405–416.

Andreas Jungherr, Pascal Jürgens, and Harald Schoen. 2012. Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, to, sander, pg, & welpe, im predicting elections with twitter: What 140 characters reveal about political sentiment. *Social Science Computer Review*, 30(2):229–234.

Marcus Kreuzer and Vello Pettai. 2003. Patterns of political instability: Affiliation patterns of politicians and voters in post-communist estonia, latvia, and lithuania. *Studies in Comparative International Development*, 38(2):76–98.

Pulkit Mehndiratta, Shelly Sachdeva, Pankaj Sachdeva, and Yatin Sehgal. 2014. Elections again, twitter may help!!! a large scale study for predicting election results using twitter. In *Big Data Analytics*, pages 133–144. Springer.

Saif M Mohammad, Xiaodan Zhu, Svetlana Kiritchenko, and Joel Martin. 2014. Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management*.

Mikael Persson, Anders Sundell, and Richard Öhrvall. 2014. Does election day weather affect voter turnout? evidence from swedish elections. *Electoral Studies*, 33:335–342.

Saatviga Sudhahar, Giuseppe A Veltri, and Nello Cristianini. 2015. Automated analysis of the us presidential elections using big data and network analysis. *Big Data & Society*, 2(1):2053951715572916.