Analyses of Methods for Prediction of Elections Using Software Systems

Lavdim Beqiri[a*], Zoran Zdravev[b], Majlinda Fetaji[c], Bekim Fetaji[d]

[a]*University Goce Delcev Shtip /Computer Science, Shtip, North Macedonia*
[b]*South East European University/Computer Science, Tetova, North Macedonia*
[c]*Mother Teresa University/Computer Science, Skopje, North Macedonia*
[a]*Email: Lavdim.beqiri@gmail.com,* [b]*Email: Zoran.zdravev@ugd.edu.mk,* [c]*Email: m.fetaji@seeu.edu.mk,*
[d]*Email: bekim.fetaji@unt.edu.mk*

Abstract

The primary objective of this research study is to review and analyze the published literature regarding the possibilities of forecasting and predicting the result of elections using software systems. The factors motivating research institutions and individuals to consider research impact on prediction of elections are manifold. Understanding the impact of different software tools, algorithms and social networking software applications on prediction of elections is a vital, and often overlooked, element of forecasting the election results. The literature review was conducted to examine methods and current software applications and practices as well as projects on election predictions. The review focused in particular on social media applications and different methods on accessing the opinion of the potential voters. The review draws on an international literature, although it is limited to English language publications. The findings identify the different methods used, the advantages and disadvantages of different approaches and the methods that are used currently and that have shown most effective results and recommendations are provided.

Keywords: machine learning; social networks; election prediction; methods; twitter.

1. Introduction

orders have been surveying populaces to attempt to follow the notoriety of races up-and-comers and individuals from governments.

----

* Corresponding author.

Those are frequently led by methods for telephone, on the web or even face to face overviews, which can be extremely tedious and generally experience the ill effects of restricted example sizes and predisposition. Social Media has brought us closer than even before and has provided a common platform for us to communicate with one another [1]. We live in a world where everyone is connected, in one way or another. Almost 4 billion people in the world are active on social media right now. With the fast development of online social stages, for example, Facebook or Twitter, any individual can now openly communicate their perspectives and suppositions, adding to an ever-growing pool of straightforwardly available information. This has open the entryway for another road of examination, that looks to utilize this valuable asset to figure surveys and political race results without reviewing the populace [13,14]. The nature of forecasts traverses a somewhat wide reach and various voices have communicated worries over these techniques, contending that there are numerous components impacting everything that may change their dependability [2]. In spite of the fact that anticipating has been utilized ordinarily in various fields, it has a short history in political theory. Anticipating political occasions began in the last part of the 1970s when [7] explored the impact of the monetary condition in the political race year also the officeholder parties in a determining model. Reference [18] analyzed the connection between the aftereffects of a political decision with the past ones. Reference [5] built up a model utilizing the president's work endorsement and a monetary factor as free factors. Reference [3] added a period subordinate variable to improve the presentation of the guaging model. The needy variable of the model was the level of the occupant party votes, and the free factors were GDP development, the officeholder president's work endorsement rating in June of the political race year, and the back-to-back terms that the occupant party oversees the nation [8]. He utilized the normal least-squares (OLS) technique to gauge the boundaries of the straight relapse model. Afterward, Reference [10] used his model, which is called "Time for change determining model", to figure out political decision.

2. Research Methodology

The review used a number of systematic review methods, including identifying a question, developing a well-defined search strategy, using pre-defined select criteria and completing data extraction sheets for each study [12]. The main advantage of this approach is that it seeks to implement standard strategy that can be easily be repeatable from other researchers in their analyses and draw the same conclusions.

3. Analyses Conceptual Search Strategy

There were many sources available to and utilized by the author. Electronic searches of databases such as Google Scholar, ProQuest, and Science Direct and Synergy were completed and articles were selected. The following were used as search terms: prediction of elections using software, election forecasting, predicting elections with twitter, voters perceptions. 41 articles have been identified, from which 20 were useful from these searches. Also a manual search in the Periodicals section of Biometrika was performed. 1 article was retrieved from this search. Articles gathered ranged in dates from 2001-2018. It was necessary for the author to look as far back as 2000 as there was not a lot of literature regarding prediction of elections using software found during the searches. Of the 11 articles found 10 were quantitative studies and 1 was a literature review. The literature spanned the globe coming from Italy, USA, Germany Australia, Sweden, Britain, Canada, and Nigeria. The

following themes emerged from the literature: social media; machine learning algorithms, clustering, twitter and sentiment analyses [16]. The literature study represents a review of these themes.

4. Background Research

A.   Machine Learning

Machine Learning MI has a place with the field of Artificial Intelligence, which thusly is the field of insight that can be exhibited by machines. The objective of Machine Learning is regularly supposed to be to figure out how to decide, or forecasts " without being unequivocally customized to play out the task"[6].

To do this the investigation of Machine Learning is vigorously centred around finding either calculations or numerical or factual models that can be utilized to prepare the Machine being referred to [4]. For the most part Machine Learning is partitioned into the classifications of regulated and unaided realizing, where managed learning, just stated, takes a bunch of information sources and known classifications and figures out how to characterize new contributions to these classifications, and solo learning takes a gander at information and discovers structure and examples in that information without understanding what that structure may be [11].

B.   Clustering

The task of clustering is to take a gander at a bunch of information and arrange that information into sets, or gatherings, that are considered to be more comparable than different gatherings. Since it isn't known heretofore what gatherings will be found, concerning how grouping is utilized inside Machine Learning, it falls into the solo class [4]. Grouping isn't a calculation, yet the assignment to be tackled. Various sorts of calculations are utilized to perform bunching and are worked in various things. Sorts of grouping models incorporate yet are not restricted to network models, for example, various leveled bunching that group dependent on distance, centroid models, for example, k-implies that group dependent on mean, diagram based models, for example, HCS that group in chart structure [7].

5. Twitter as A Basis for Research

Several researchers in their studies [4,9,15,19,20] did a writing audit of the past examination that had been done on Twitter as a reason for political decision expectation. Subsequent to experiencing this exploration, they arrived at the resolution that nations where the web client rate is above 80% are for the most part fit for investigation utilizing Twitter as the base for political decision expectation. In the present, the online media network assumes a huge part in dividing data among people. This joins data about news and occasions that are by and by happening in the around the world. Envisioning political race results is as of now transforming into an interesting examination subject through online media. In this article, we proposed a procedure to envision political race results by combining sub-occasion revelation and nostalgic examination in miniature websites to separate just as envision political tendencies revealed by those online media clients. Online media has throughout the years halfway become a stage to communicate suppositions and examine recent developments. Inside the field of Computer Science, Twitter has been utilized both as the reason for political investigation - for

instance utilizing supposition examination to anticipate political race results - and inside the field of group examination, where the subject of how to best plan and utilize a calculation to extricate points from tweets has been considered. Solution for the hierarchical location classification of Twitter users has been implemented using Python as the programming language. Using Python programming language to build a model that predicts the election result by getting the required data from Tweepy and by importing powerful python libraries like Numpy for all the athematic and multi-dimensional array usage and Matplotlib for convenient graphical representations.

6. Review of Methods And Research Projects

A number of various research projects have been realised in on applying machine learning calculations to Twitter information to estimate assessment of public sentiment results or political decision results. Some of them have been reviewed and allude the intrigued peruser to [17] for more in-depth surveys of the writing. A pioneer work around there was that of [14] whose model accomplished a mean normal mistake (MAE) of 1.65% when anticipating consequences of the 2009 German government political race [9]. Creators utilized Twitter notice considers an immediate marker of a competitor's notoriety, a technique that has been considered by a few different functions too, regularly in mix with an estimation investigation of tweets content [13]. Specifically [9] accomplished 90% exactness in anticipating the main two up-and-comers in different regions during Brazilian civil decisions, and Saleiro and his colleagues [12] accomplished a MAE of 0.63% when attempting to anticipate assessment of public sentiment results during the bailout (2011-2014). Twitter has pulled in the consideration of a few analysts in territories like anticipating the customer brands, majority rule appointive occasions, film industry, prevalence of VIPs, the securities exchange, and so forth Conclusion investigation over a Twitter-based informal organization offers a quick and proficient method of checking the public assumption. A few explorations contemplate examined the assessment expectation task over Twitter utilizing AI procedures, with the thought of Twitter-explicit informal community structure, for example, retweet. They additionally focus on finding both immediate and stretched out terms identified with the occasion and in this manner understanding its impact. Mostly utilized directed AI strategies [13], for example, uphold vector machines (SVM), Naive Bayes, greatest entropy and fake neural organizations to order the Twitter information utilizing unigram, bigram and unigram + bigram (half and half) highlight extraction model for the contextual analysis of US Presidential Elections 2012. Further, they consolidated the aftereffects of notion investigation with the impact factor created from the retweet tally to improve the expectation precision of the errand. Trial results show that SVM beats any remaining classifiers with greatest exactness of 88 % in foreseeing the result of US Elections 2012, and 68 % for Indian State Assembly Elections 2013.

7. Use of Sentiment Analysis

The most important and striking works concerning political race expectation with respect to Twitter are incompletely founded on slant investigation - Andranik and his colleagues [14] arrived at the resolution that, utilizing estimation investigation on twitter information, political decision expectations can be made, while O'Connor and his colleagues [15] utilized comparative techniques to reason that tweets, in any event, could be utilized instead of or as an enhancement to customary surveying strategies. Reference [16] examined the

conceivable outcomes of utilizing conclusion examination along with relapse investigation on the Indian decisions and inferred that the greatest test in a nation, for example, India as the assortment of information. Assumption examination is 'the errand of recognizing positive and negative assessments, feelings, and assessments' [17]. Since its start, feeling examination has been subject of a serious exploration exertion and has been effectively applied to different territories. A few models incorporate helping clients in their improvement by giving them intriguing and strong substance [19], anticipating the result of a political race [1], film deals [18] and item audit estimations. The scope of supposition examination strategies changes from distinguishing extremity (positive or negative) to a complex computational treatment of subjectivity, assessment and opinion [20]. Specifically, the examination on feeling extremity investigation has brought about various developed and openly accessible instruments (paid just as free, for example, SentiStrength [17], Alchemy, LingPipe, ElasticSearch slant analyzer, Lexalytics, Recursive Neural Tensor Network [20], DatumBox, text-handling, GATE and NLTK as discussed by [18].

8. Using Other Methods

A few other research studies have utilized different techniques and have been made attempting to anticipate political race results with changing achievement, see for instance [15], who utilized relapse investigation to some achievement, however none explicitly utilizing theme demonstrating. Nearest comes [20], whose review somewhat utilized multinomial theme displaying along with network investigation to anticipate the 2011 Nigerian presidential elections and arrived at the resolution that the method could be utilized to distinguish content-based organizations however didn't in an agreeable manner foresee the political decision result [15].

Table 1: Important decision variables [6]

Key variable	Description	Mean Gini index score
Hmix	Household Mix	838.3
Party	Party of Registration	708.9
Cd	Congressional District	319.7
Race	Race	233.7
vh_score	Voting History Score (or how many times has someone voted)	188.4
sex	Gender/Sex	130.8
vh08g	Voted in previous General Election (0 or 1)?	124.2

Table 2: Model predictions versus actual results [8]

Election Republican, %	Democratic, %	Difference from actual result	%
FiveThirtyEight analysis (December 5, 2014)	57.8	42.2	1.9
CVS model predicted (December 5, 2014)	56.1	43.9	0.2
Actual result	55.9	44.1	-

9. Conclusions

Although Social Media has its downsides, we absolutely cannot deny the potential it holds in various domains. From businesses to even politics, social media plays a pivotal role. Social media analytics is the practice of gathering data from social media websites and analyzing that data using social media analytics tools to make decisions. The most common use of social media analytics is to mine customer sentiment. Multiple social media platforms have been created and many of them have been discontinued but Twitter still rides strongly with its large following. Twitter provides a valuable tool for social media analysis because it is a never-ending resource which is self-sufficient and keeps along with the times. Using machine learning together with implementing the third part of the algorithm to see if it is possible to collect tweets from running parties and run them through the algorithm to predict the winner in that state based on what topics the tweets should be classified into.

References

[1]. Gayo-Avello, D.: No, you cannot predict elections with twitter. IEEE Internet Computing 16(6), 91{94 (2012). doi:10.1109/MIC.2012.137

[2]. Garcia, A.C.B., Silva, W., Correia, L.: The PredNews forecasting model. In: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age. Association for Computing Machinery, New York, NY, USA (2018). doi:10.1145/3209281.3209295

[3]. Gayo-Avello, D.: Analyzes predicting elections with twitter. IEEE Internet Computing 16(5), (2012). doi:10.1109/MIC.2012.142

[4]. Gayo-Avello, D.: A meta-analysis of state-of-the-art electoral prediction from twitter data. Soc. Sci. Comput. Rev. 31(6), 649{679 (2013). doi:10.1177/0894439313493979

[5]. Grimaldi, D., Cely, J.D., Arboleda, H.: Inferring the votes in a new political landscape: the case of the 2019 Spanish presidential elections. J Big Data 7 (2020). doi:10.1186/s40537-020-00334-5

[6]. Hassin, Y., Peleg, D.: Distributed probabilistic polling and applications to proportionate agreement. Inf. Comput. 171(2), 248{268 (2002). doi:10.1006/inco.2001.3088

[7]. Holley, R.A., Liggett, T.M.: Ergodic theorems for weakly interacting infinite systems and the voter model. Ann. Probab. 3(4), 643{663 (1975). doi:10.1214/aop/1176996306

[8]. Huberty, M.E.: Multi-cycle forecasting of congressional elections with social media. In: Proceedings of the 2nd Workshop on Politics, Elections and Data. PLEAD '13, pp. 23{30. Association for Computing Machinery, New York, NY, USA (2013). doi:10.1145/2508436.2508439

[9]. Jungherr, A., J¨urgens, P., Schoen, H.: Why the pirate party won the Brazilian election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, t. o., sander, p. g., & welpe, i. m. \predicting elections with twitter: What 140 characters reveal about political sentiment". Soc. Sci. Comput. Rev. 30(2), 229{234 (2012). doi:10.1177/0894439311404119

[10]. Jungherr, A., Schoen, H., Posegga, O., J¨urgens, P.: Digital trace data in the study of public opinion: An indicator of attention toward politics rather than political support. Social Science Computer Review 35(3), 336{356 (2017). doi:10.1177/0894439316631043

[11]. Klamser, P.P., Wiedermann, M., Donges, J.F., Donner, R.V.: Zealotry effects on opinion dynamics in the adaptive voter model. Phys. Rev. E 96 (2017). doi:10.1103/PhysRevE.96.052315

[12]. Jungherr, A., Schoen, H., Posegga, O., J¨urgens, P.: Digital trace data in the study of public opinion: An indicator of attention toward politics rather than political support. Social Science Computer Review 35(3), 336{356 (2017). doi:10.1177/0894439316631043

[13]. Mobilia, M.: Does a single zealot affect an infinite group of voters? Phys. Rev. Lett. 91, 028701 (2003). doi:10.1103/PhysRevLett.91.028701

[14]. Audickas, L., Cracknell, R., Loft, P.: UK Election Statistics: 1918-2019 { A Century of Elections (2020). https://commonslibrary.parliament.uk/research-briefings/cbp-7529/

[15]. Bovet, A., Morone, F., Makse, H.A.: Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. Sci. Rep. 8, 8673 (2018). doi:10.1038/s41598-018-26951-y

[16]. Caldarelli, G., Chessa, A., Pammolli, F., Pompa, G., Puliga, M., Riccaboni, M., Riotta, G.: A multi-level geographical study of italian political elections from twitter data. PLoS One 9(5), 1{11 (2014). doi:10.1371/journal.pone.0095809

[17]. Chen, P., Redner, S.: Majority rule dynamics in finite dimensions. Phys. Rev. E 71 (2005). doi:10.1103/PhysRevE.71.036101

[18]. Clifford, P., Sudbury, A.: A model for spatial conflict. Biometrika 60(3), 581{588 (1973). doi:10.1093/biomet/60.3.581

[19]. Dokoohaki, N., Zikou, F., Gillblad, D., Matskin, M.: Predicting swedish elections with twitter: A case for stochastic link structure analysis. In: 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 1269{1276 (2015). doi:10.1145/2808797.2808915

[20]. Fernley, J., Ortgiese, M.: Voter models on subcritical inhomogeneous random graphs (2019). 1911.13187 8. Fink, C., Bos, N., Perrone, A., Liu, E., Kopecky, J.: Twitter, public opinion, and the 2011 Nigerian presidential election. In: 2013 International Conference on Social Computing, pp. 311{320 (2013). doi:10.1109/SocialCom.2013.50