International Journal of Scientific Advances
ISSN: 2708-7972
Volume: 2 | Issue: 3 | May - Jun 2021 Available Online: www.ijscia.com
DOI: 10.51542/ijscia.v2i3.34

Devising an Algorithm for Election Prediction Using Survey of Voters Opinions

Lavdim Beqiri[1], Zoran Zdravev[2], Majlinda Fetaji[3] and Bekim Fetaji[4*]

[1,2]Department of Informatics
University Goce Delcev, Shtip, North Macedonia

[3]Department of Computer Sciences,
South East European University, Tetova, North Macedonia

[4]Department of Informatics
Mother Teresa University, Skopje, North Macedonia

E-mail: Lavdim.beqiri@gmail.com; Zoran.Zdravev@ugd.edu.mk;
m.fetaji@seeu.edu.mk; bekim.fetaji@unt.edu.mk

*Corresponding author details: Bekim Fetaji; bekim.fetaji@unt.edu.mk

ABSTRACT
The purpose of this research study is to analyze how we use voter polls to predict elections and to design an algorithm to predict elections. We propose a method of prediction based on learning algorithm to determine the political profile of a voter group by obtaining a linear hierarchy on the attributes that weights the number of instances that are more relevant. Our process starts with opinion survey collected directly from the target group of voters. Having a linear attribute hierarchy that expresses the political preferences of voters allows the application of a holistic approach to distribute the potential number of votes among the parties involved. We applied our electoral outlook model in the Kosovo election case study in from February 2021. The devised algorithmic model may also be applied to other situations. Data analysis not only provides new analysis opportunities, but also faces many challenges. In our case, we listed the limitations of the research. The research attempts to promote the implementation of the algorithm by extending the processing of the information generated by the learning algorithm to improve the prediction of elections and winning parties. Discussed of all data analysis challenges, and present, discuss, and argue insights.

Keywords: data analytics; learning algorithm; recognizing political profiles; comparative analyses

INTRODUCTION

Data analysis and machine learning from AI, according to [1] can now be used in a variety of ways, from mixing and developing data analysis, combining, mixing, merging and creating data analysis tools from different data sources, to developing new data models for data analysis, data modeling or data visualization. Data analytics and machine learning from machine learning and artificial intelligence (AI) can now use a range of traditional and modern methods to mix many and many different data sources and develop new data models for data analysis, data mining and data visualization. [2]

There is an urgent need to develop and integrate new mathematical, visualization and computational models with the ability to analyze data to retrieve useful information. By understanding the value of data and data analysis, computer scientists and political analysts can learn from each other. Data derived from voters and citizens must also be analyzed to be actionable.

One of the common decision-making processes that has become the subject of many studies is "voting" in a democratic system. Voting is a process that describes how the preferences of a group of people determine collective decision-making. The selection of representatives is a very uncertain process.

In addition, the Uniform Random Sampling Procedure has been widely used as a means of electoral perspectives through public opinion polls. The size of the sample depends mainly on the variance of the population: since the population of events to be estimated is more diverse, the sample size will increase regardless of the population size of the population [5].

Random sampling as discussed by [6] also requires increasing the sample size at an exponential rate relative to the probability of success of the sampled event. In other words, identifying events with high uncertainty requires a much larger sample than events with lower uncertainty.

For example, if there is a need to identify election trends and voters have greater uncertainty to determine their preference for only competing candidates, then in addition to increasing the sample size, the uncertainty of the respondents' preferences should also be considered Sex., Which may cause conflicting or inaccurate answers to survey questions.

The use of statistical methods has always been the main tool for conducting election surveys, but its effectiveness for this survey task is unclear, because the actors involved abuse this tool to disseminate results at their convenience. In addition, as discussed by [7] sampling can be improved by considering other knowledge about the population to be sampled.

In order to gain knowledge about voters, we recommend establishing a political profile of voters from the initial poll. To this end, we apply data extraction techniques to the respondents' responses to characterize the regions that support and oppose each candidate. This, in turn, means designing and implementing more specific polls that match voters' needs with their political preferences.

The learning algorithm has become a commonly used predictive model in the field of artificial intelligence. In this article we present the application of learning algorithms as a technique to extract information from public opinion polls. Therefore, our main contribution is set to create an innovative algorithm that extends the processing of the information generated by the learning algorithm to improve the analyses and prediction of elections and winning parties.

In this case, the learning process begins with the collection of data from multiple sources through multiple surveys. The next step is data preparation, which means that data preprocessing methods can solve data-related problems and reduce spatial dimensions by eliminating unnecessary data. Since the amount of data used for learning remains enormous, system decision-making is problematic. In this case, the algorithm uses logic, probability, statistics, and certain control theories to analyze the data and retrieve it from the initial experience.

Due to the lack of analytical tools to simulate voting trends, we have developed a new method to identify the political profile of groups of voters. A key element of our method is to identify a set of "primary attributes" that characterize the political behavior of voters.

We recommend the use of polls as a preliminary method to identify the relationship between voters' political concerns and preferences. We use the Internet as a medium to promote and collect these surveys. We call a profile to a set of primary attributes of the voters, which infer the value of the class (political decision). Our proposal includes the application of learning algorithms to automatically identify the personal data that characterizes the political preferences of voters. The parties attempted to identify and modify voters 'personal data in different ways, formulating different political plans based on the voters' responses to these plans.

LITERATURE REVIEW

In order to identify articles that employ data analytics and its application in prediction of elections using election pools and surveys, extensive efforts were made. Several databases were searched: the extensively the IEEE digital library, ACM digital library, and Google scholar, containing more than 1 million journal articles, conference papers, and other publications on computer science.

An opinion poll, in its traditional elaboration form, usually reflects outlying questions about candidates, and about the political competition, such as: popularity indexes, perceptions on the nature of the candidates or their images, the impact of their campaigns, etc. Often the factors that are measured through those surveys point more to the interest of the candidates or their parties, than to the interest or perception of voters.

The literature on electoral projections is basic, because studies should nourish it, as in the case of statistical analysis, which are scarce [6]. The lack of specialized bibliography, is due to the fact that since 1993, are disseminated by the Federal Electoral Institute IFE, and the state electoral bodies, the overall results and with some levels of disaggregation; what has meant is that there are no historical series of voting, nor criteria to build units of comparison.

The prediction of the 2009 elections in Germany, using the techniques referenced, as shown in [3] why the party won the German election of 2009 or the trouble with predictions: was done by taking into account the frequency of mentions and to get the total mentions, replication of mentions and percentages of mentions. The sample is less than a month, and takes representative days. It also takes into account the progression of the followers. The analysis of the results is quantitative.

The data mining aims according to [1] uncover patterns and relationships to decision supporting system and making predictions. First, the classification of the data, by a process of unsupervised learning as the grouping clustering, brings the find of groups that are different but the individuals are equal among themselves as noted [7] In the methodology of data mining.

We select the use of the data mining software called RapidMiner, because it is an easy tool, and where different jobs are choosing: like, Empirical studies of machine learning based approach for opinion mining in tweets, [5] The comparison of different classification techniques have been discussed and followed guidelines by [6].

METHODOLOGY

It is important to study the relationship between the historical trend of the vote and the electoral results of a specific process; it is important because it allows us to make predictions, which can, in good measure, sensitizing political actors and citizens about the possible results of the electoral process.

It is pertinent to note, that the research was carried out, by ordering the results of the processes of parliament elections of the state of Kosovo in 2021, to develop series of votes, which were necessary to carry out the projections. The election results are not entirely accidental events, fully relieved by past events, and that much of what happens in local processes allows us to contemplate the possible scenarios of the local process.

Thus, in the case of the local executive, it incorporates data from the last elections for 2021, data were analyzed from parliament elections, due to the difficulties to normalize the data and the lack of the data themselves; it was determined to use the data for parliament elections.

ELECTORAL SURVEYING

Based on the weights of factors discussed by [4] and applying the procedure for vote distribution, the expected percentages of votes for the candidates are obtained and presented in Figure 1 below. This figure includes the percentage of expected votes per party.

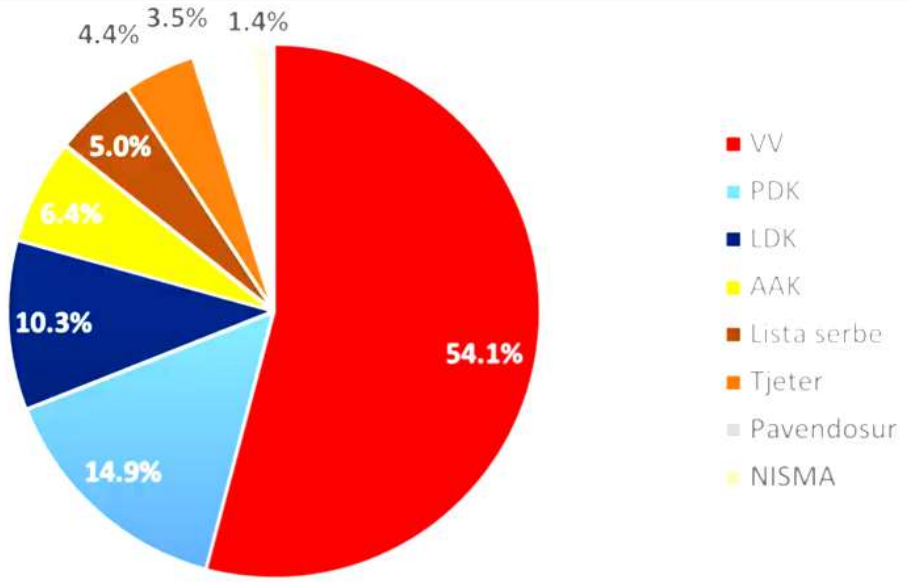

FIGURE 1: Survey expected percentage of votes from elections

Although in the basic statistics, parties have obtained a same number of favorable votes, the application of our model provided additional information that statistical methods cannot assess. For example, the profile of the followers for one of the candidates is more homogeneous because he obtains a higher positive weight (preferred candidate) and a less negative weight (non-preferred candidate) than the values obtained for the other candidates. This last result was confirmed with the actual values obtained in the election, where the candidate with the more homogeneous followers won the contest.

In elections held on 14 February 2021, the political party "Vetëvendosje" won the election with 50.28% in total votes and ahead of his closest contender party "Democratic Party of Kosovo" with 33.27% of total votes. While the results of our model: Decision Trees + Heuristics to build a hierarchy of attributes + Distribution of Votes, obtained a difference of 13.71 points in favour of Vetevendosja, the difference between actual and our result is 3.82% percentage points, being acceptable for a computer model that predicted the expected results of the election, four days ahead of the election date

In the Table 1 below are the actual results from the elections in Kosovo.

TABLE 1: The actual results from the elections in Kosovo

Party	Votes	%	Seats	+/–
Vetëvendosje	438,335	50.28	58	+29
Democratic Party of Kosovo	148,285	17.01	19	–5
Democratic League of Kosovo	110,985	12.73	15	–15
Alliance for the Future of Kosovo	62,111	7.12	8	–5
Serb List	44,407	5.09	10	0
Social Democratic Initiative	21,997	2.52	0	–4
Turkish Democratic Party of Kosovo	6,496	0.75	2	0
Vakat Coalition	5,366	0.62	1	–1
New Democratic Initiative of Kosovo	3,305	0.38	1	0
Romani Initiative	3,172	0.36	1	New
New Democratic Party	2,885	0.33	1	0
Social Democratic Union	2,549	0.29	1	New
Egyptian Liberal Party	2,430	0.28	0	–1
United Community	2,217	0.25	0	New
Unique Gorani Party	2,161	0.25	1	0
Ashkali Party for Integration	2,138	0.25	1	0
Democratic Ashkali Party of Kosovo	1,960	0.22	0	0
Civic Initiatives for Freedom, Justice and Survival	1,508	0.17	0	New
Our Initiative	1,375	0.16	0	New

	Movement for Integration	1,261	0.14	0	New
	Innovative Turkish Movement Party	1,243	0.14	0	New
	Progressive Movement of Kosovar Roma	1,208	0.14	1	New
	Fjala	1,087	0.12	0	0
	United Roma Party of Kosovo	1,074	0.12	0	–1
	Coalition Together (GIG–PG)	1,010	0.12	0	0
	Kosovar New Romani Party	600	0.07	0	0
	Serbian Democratic Alliance	476	0.05	0	New
	Albanian National Front Party	155	0.02	0	New
Total		**871,796**	**100.00**	**120**	**0**
Valid votes		871,796	96.38		
Invalid/blank votes		32,756	3.62		
Total votes		**904,552**	**100.00**		
Registered voters/turnout		1,851,927	48.84		
Reference Source: CEC, CEC, CEC					

Table 1 includes the results taken from the referenced official web pages (in Albanian). This chart details the candidate coalitions or parties in x axis, and the number of votes in the y axis.

The first lefmost bar is labelled for the party coalition that supported the parties. Note also that the expected number of null votes predicted by the model coincided with the actual proportion (fourth bar), i.e. was higher.

THE ALGORITHM

We recommend using learning algorithm (such as the algorithm designed below) to determine the main attributes that characterize the political preferences of the population. One limitation of the algorithm is that the instance must only have discrete values associated with each attribute. If the response matrix contains significant noise, the classification rules may include a high error rate. To determine the attribute that provides the maximum information gain, the entropy in the training matrix can be calculated as:

$$Ent = -\Sigma c(f(a) \cdot \log 2f(a)) \quad (1)$$

Entropy is the average value of the amount of information required to classify an object, where f (a) is the probability that such an object corresponds to a specific value of class "a". Also, the entropy of each attribute is calculated based on the number of instances with different attribute values. The information transmitted by each attribute A is calculated as:

$$Inf(A) = -\Sigma v(f(v) \cdot (\Sigma cf(v) \cdot \log 2f(v))) \quad (2)$$

where f (a) is the conditional probability to obtain the value of class a, given that the attribute A has the value v and the average of the gain of an attribute A (the entropy of the attribute A), is computed as:

$$Ent(A) = -\Sigma c(f(v) \cdot \log 2f(v)) \quad (3)$$

The information gain is computed for each attribute as:

$$ttain(A) = (Ent - Inf(A))/(Ent(A)) \quad (4)$$

the attribute A with the highest information gain is selected as the split node in the current tree, so that a branch A of A is formed for each value in the domain, and then A is discarded from the attribute set. Then apply the learning algorithm repeatedly to continue to form a complete decision tree. However, considering, for example, voters opposed to each side, different trees were built.

CONCLUSION

The main purpose of the research study was to investigate data analytics and its applications in prediction of elections by devising an algorithm. Primarily the focus was on devising an algorithm on predicting elections and winning political parties based on survey of voters. In conclusion, the survey done on various models specific to the area of political parties and prediction of the results and winning party, is clear that data science has evolved well in predictive analysis with regard to prediction and visualisation. We have shown here that learning algorithms can be applied to determine the most relevant attributes characterizing the political preference of a population (called the profiles of the voters), in favour and against each of the contending candidates in a democratic election. The set of profiles is automatically generated after processing the opinion polls via learning algorithms.

We proposed a new method of prediction based on learning algorithm to obtain a linear hierarchy on the attributes that appear in a decision tree. The attributes hierarchy weights the number of instances which have chosen a specific class value, and whose logical rules (paths) passes through the attributes in question.

So, if one wishes to characterize political preferences of a population, it is important to determine which attributes are more relevant than other. But this hierarchy has to balance the attributes according to whether there are more (or less) voters who use that attribute to make its political decision.

In the election of our case study, the contest was won by the candidate with 9.05 percentage points above its nearest competitor. The difference between reality and our model was 4.66 percentage points, whereas traditional statistical methods had obtained a tie between the two leading candidates.

The results of our model represent an accept- able outcome for a computer model that predicted the expected results of the election, four days prior to the election date. Our approach has implication for political science, where it can support analysis.

Recommendations

As such and according to the results of the study, some recommendations can be given:

The data analysts must raise their seriousness in developing secure and serious data analytics applications. The data science analyses applications should be user friendly and usable, so that will increase the clients' satisfaction.

By having serious dissemination and presentation of the data science application and providing sufficient training, can raise the confidentiality towards the use of the data science software tools.

We foresee that our methodology can also be applied to model product advertising campaigns, where from customer surveys, it can characterize the product of interest, in addition to contrast such product against other competing advertised goods.

REFERENCE

[1] Burstein, Frada and Holsapple, 2018, CW., Handbook on Decision Support Systems 1: Basic Themes, International Handbooks on Information Systems, Springer, 2018.

[2] Gloor, P. A. and Krauss, J. and Nann, S. and Fischbach, K. and Schoder, D., 2019, Web Science 2.0: Identifying Trends through Semantic Social Network Analysis, ICCSE, 2019.

[3] Hans Ulrich Buhl, 2011, From Revolution to Participation: SocialMedia and the Democratic Decision-Making Process, BISE Editorial, 2011.

[4] Md Safiullah, et al, 2017, "Social media as an upcoming tool for political marketing effectiveness" (2017). Elsevier 2017.

[5] Shira Fano and Debora Slanzi , Using Twitter Data to Monitor Political Campaigns and Predict Election Results. Springer international publishing AG 2017.

[6] Tsakalidis Adam, et al. , predicting elections for multiple countries using twitter and polls. IEEE 2015.

[7] Zhihan Lv, et al 2016, Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics. IEEE 2016.