

# Survey Analyses of The Specific Impacting Factors in Devising a Machine Learning Prediction model for The General Election Process in Kosovo

Lavdim Beqiri<sup>a</sup>, Majlinda Fetaji<sup>b\*</sup>, Zoran Zdravev<sup>c</sup>, Bekim Fetaji<sup>d</sup>

<sup>a,c</sup>*Department of Informatics, University Goce Delcev, Shtip 2000, North Macedonia*

<sup>b</sup>*Computer Sciences, South East European University, Tetovo, 1200, North Macedonia*

<sup>d</sup>*Informatics, Mother Teresa University, Skopje 1000, North Macedonia*

<sup>a</sup>*Email: Lavdim.beqiri@gmail.com,* <sup>b</sup>*Email: m.fetaji@seeu.edu.mk,* <sup>c</sup>*Email: Zoran.Zdravev@ugd.edu.mk*

<sup>d</sup>*Email: bekim.fetaji@unt.edu.mk*

## Abstract

The focus of the research study was analyses of impacting factors and later to incorporate those insights into variables to be measured for devising a machine learning predictive model for prognosis and prediction of the general election turnout in Kosovo. We have developed a novel method for recognizing the main impacting factors in elections. Our method shows that finding out whether different ways of collecting different data of election voters can lead to much better prediction and understanding of the election process. In order to do that we needed to analyze the specific impacting factors in the election process in Kosovo are investigated during the study. The data has derived from an originally collected survey dataset that contains the impacting factors previously identified and assessed regarding the general parliamentary elections in Kosovo has been realized. Insights and recommendation has been discussed and argued.

**Keywords:** heuristic; election turnout; prediction; general election turnout; machine learning.

## 1. Introduction

System level analyses of turnout have also used several different approaches [1]. Some of the factors affecting turnout rates are inherently systemic (e.g. electoral system) and some can be used in an aggregated form in macro level analysis as discussed by [2] (e.g. average age of potential voters in a constituency). These macro level analyses confront two important problems. The first one is the problem of reduction [3]. Countries or constituencies do not vote, individuals do. Thus, system level propositions must be reduced to the individual level to make them plausible as theoretical explanations.

---

\* Corresponding author.

In some cases, this is quite easy. For example, compulsory voting (system level factor) increases the costs of failure to vote on the individual level (failure to vote is likely to be fined) and consequently we can assume that turnouts higher on the macro level. However, this kind of reduction is not always unproblematic [3]. The second problem relates to methods. Using aggregated individual level data to make inferences concerning individual behavioristics the risk of “ecological fallacy” [4].

System level factors associated with turnout are usually connected to electoral laws or political context. The link between compulsory voting and higher turnout rates has been clearly established. Other factors explaining high turnout are the existence of a proportional electoral system, high age requirement for voting or whether the country has a bicameral or unicameral parliament [2].

Of the factors related to the political context, the perceived closeness of elections and number of parties competing for votes seem to be important [1]. These are only a few examples of possible system level hypotheses affecting turnout rates.

## **2. Background Research**

The purpose of this study is to summarize the knowledge about prediction of elections and characteristics in terms of the strength (algorithm strength) with respect to voter characteristics (political party affiliation and their regional) which are collected from their data collected during the 2021 general elections in Kosovo. Specifically, we are interested in studying the prediction of elections with our new, innovative method using Machine Learning. We begin our analysis by considering variables that tend to predict strength with respect to political party affiliation.

We summarize our findings in a table entitled, “Election Prediction. We analyzed data sets for voting patterns run on Google Trends for 2008 to 2022. We added information regarding the political ideology of voters, and we determined the prerequisites for successful elections: if there are more committed voters than disinterested voters, the politician that is supported by the majority of the voters will emerge victorious. However, from analyses of recent elections and political history of Kosovo we conclude that they are replete with complex results: the date May 11, 2008 year, marked the transition from one-party state to a multi-party one, and again from one political actor winning to the opposition party emerging victorious.

## **3. Research Methodology**

The analyses on impacting factors on election prediction has been analyzed, and insights and results have been provided. Our method shows that finding out whether different ways of collecting different data of election voters can lead to much better prediction and understanding of the election process. For this reason, we consider above mentioned categories as a relevant factor to be measured for prediction of elections. The results show that when the survey measure the above criteria’s defined as impacting factors, they tend to establish and maintain strategic relationships to provide critical information and improve the likeliness of prediction of the election outcome. This means that researchers who would like to predict elections should know how to identify the correct information prior to process it further and collect the data and therefore to actively plan an activity of

prediction and to consciously participate in forecasting process. The aim is for them to know whom to add as category in their survey based on the information shared above. Having strategic categories identified as impacting factors can also improve the results of the forecasting the proper results and outcome of the elections.

## 4. Results

### 4.1 Strategic sampling

The strategic sample consisted of 345 voters from Kosovo, different cities. Participants were selected randomly based on their age group, preferences for different political parties. Study Participants consisted of 345 voters. We conducted the experiment with those voters. There are more female participants in the experiment. It is not the purpose of the research to analyze the sample based on the characteristics of the participants. However, we believe that the presence of more female participants can strengthen the analysis because there is scientific evidence that women are voting constantly and do not change easily their preferences for voting a certain political party.

### 4.2 Data Assessment

Cross-tabulation records have been used to assess the relationships between variables. It compares two sets of data within one chart. This reveals specific insights based on the participants' responses to different questions. To evaluate each of the categories identified as impacting factors, a scale of 20 items was developed. The project followed guidelines as discussed and recommended. These questionnaires are answered using cross-tabulation, a five-step Likert-type scale, and different combination of enquires.

3. Select which political party you will vote? / Selektoni cilën parti e adhuron dhe e kishit votuar në zgjedhjet tani

334 responses

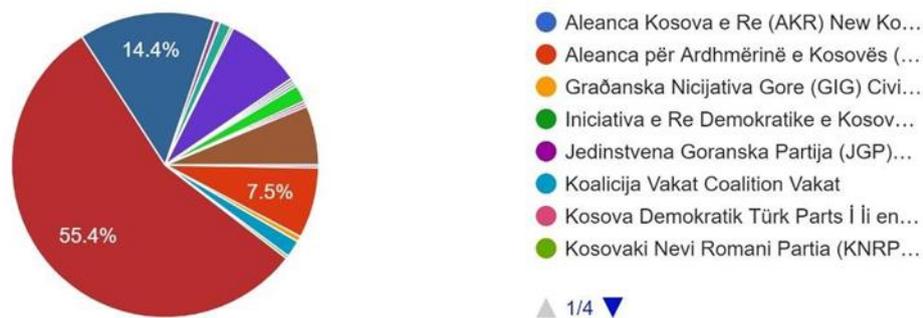


Figure 1: Enquiry for Political party preferences.

Regarding the question which political party they will vote. The results show that 55.4% would vote the Alliance for Future of Kosovo, and 14.4% would vote the Alliance for New Kosovo and 7.5% for Kosovo Turkish Democratic Party, while the remaining 22.7% for other parties.

4. Select which political party you have voted in previous elections? / Selektoni cilën parti e keni votuar në zgjedhjet e kaluara?

334 responses

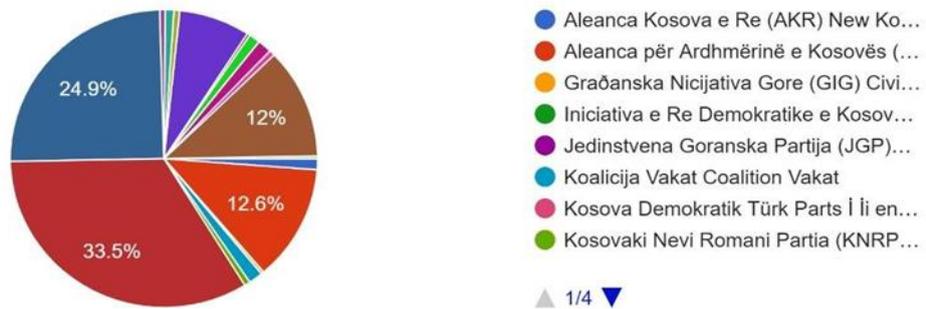


Figure 2: Enquiry for Political party voted previously.

Regarding the question which political party they have voted the previous elections. The results show that 33.5% voted the Alliance for Future of Kosovo, and 24.9% voted the Alliance for New Kosovo and 12.6% voted for Kosovo Turkish Democratic Party, and 12% did not voted at all. The remaining 17% voted for other parties.

5. Are you politically active? / A jeni aktivist partiak apo vetëm adhurues të ndonje partie?

334 responses



Figure 3: Political activity.

6. Are you married? / Gjendja e Juaj martesore

334 responses

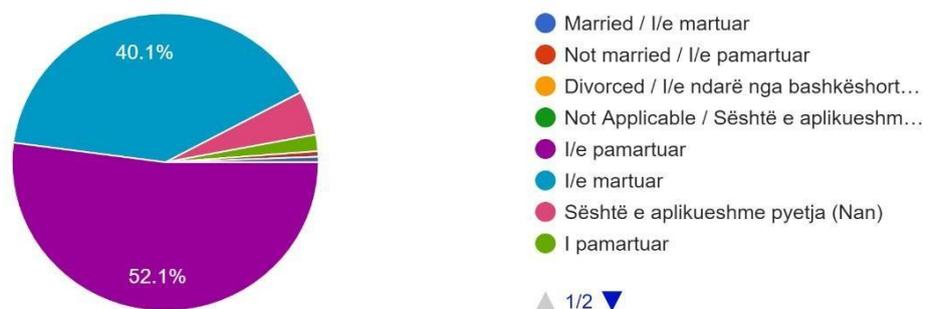


Figure 4: Marital status.

7. How many children do you have? / Sa fëmijë keni? (shkruani vetem numër)

334 responses

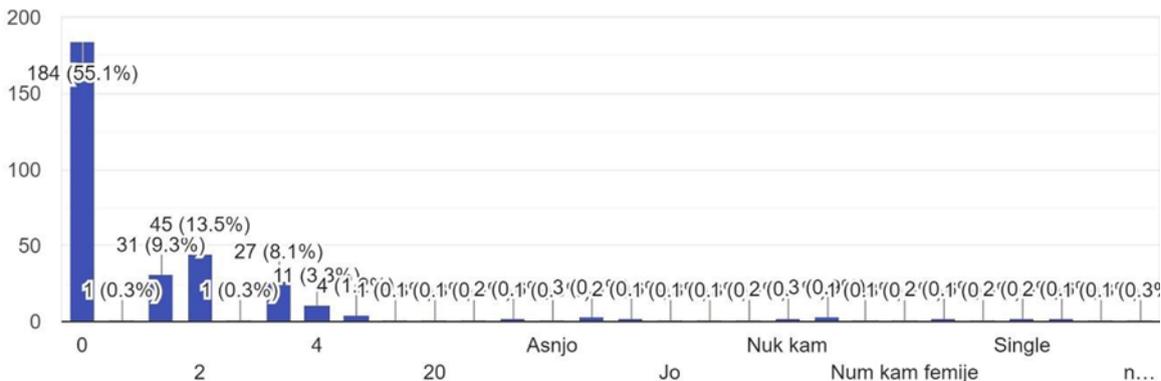


Figure 5: Enquiry for childrens.

8. Do you live in community together with other family members? / Jetoni ne komunitet se bashku me sa anetare te familjes

334 responses

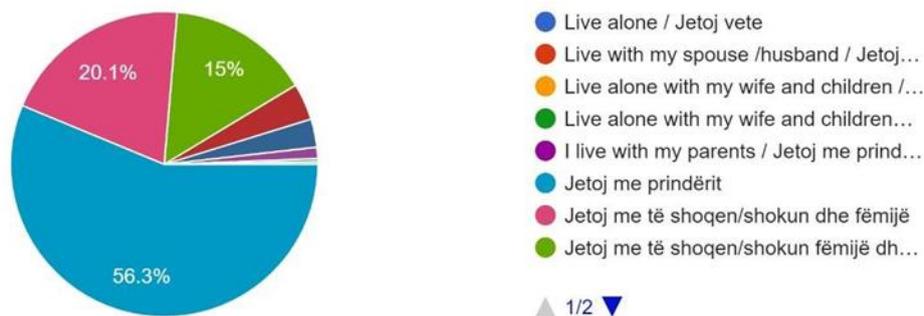


Figure 5: Enquiry for family living.

9. How many family members do you live in one house- apartment? / Sa anëtarë të familjes jeni gjithsej që jetoni së bashku në një vend (banesë, shtëpi). Shkruani vetëm numër.

334 responses

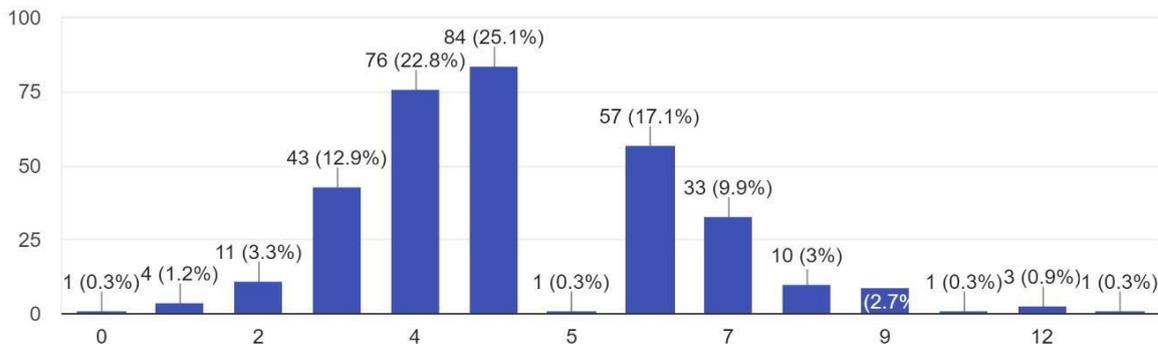


Figure 6: Enquiry how many family members live together.

Regarding the question how many family members live in one household. The results show that 55.4% would vote the Alliance for Future of Kosovo, and 14.4% would vote the agree while 9.1% are neutral. Nobody disagrees. In total larger number 90.9 % of the respondents agree that the pandemic has highly influenced the students. According to the focus group of students, especially the issues with some of teachers that do not teach but only put online materials and assessment has deviated largely and timing is shorted and this put the students in very difficult position to teach themselves and try to answer exam questions in not realistic times that their Professors could not do.

11. Select your level of education? / Ju lutemi selektoni edukimin e Juaj

334 responses

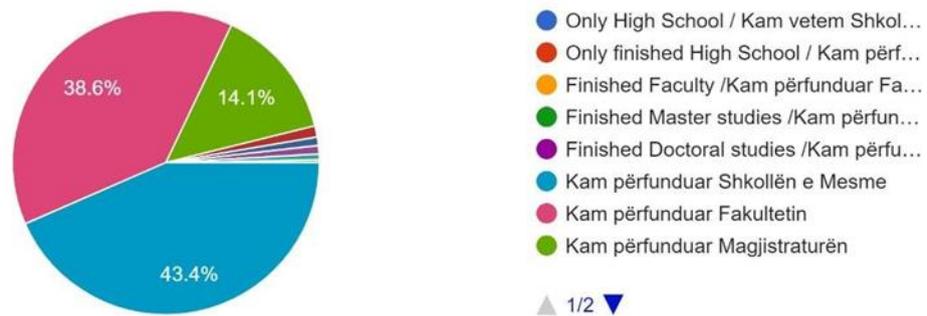


Figure 7: Education status.

12. Please select you municipality where you vote? / Ju lutemi selektoni Komunën e Juaj ku votoni?

334 responses

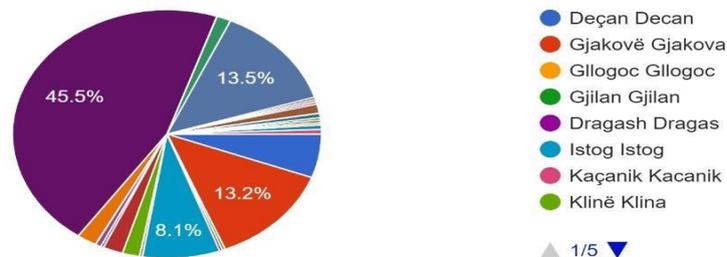


Figure 8: Enquiry for municipality they are voting.

13. Are you employed in public or private sector? A jeni të punësuar në sektorin privat apo publik?

334 responses



Figure 9: Employment analyses.

15. Do you rent anything monthly ? / A jepni diçka me qira (keni të hyra nga qiraja)?  
334 responses

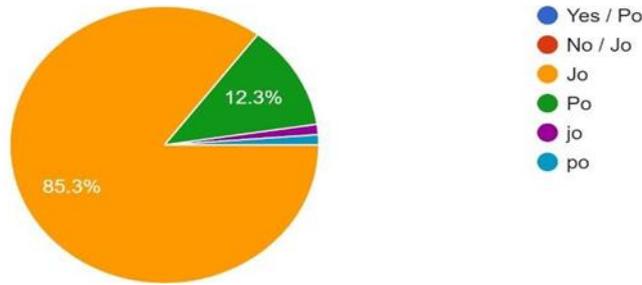


Figure 10: Renting Enquiry.

16. Are you for one ditrict of voting? / A jeni për zgjedhje direkte në 1 zonë me lista të hapura?  
334 responses

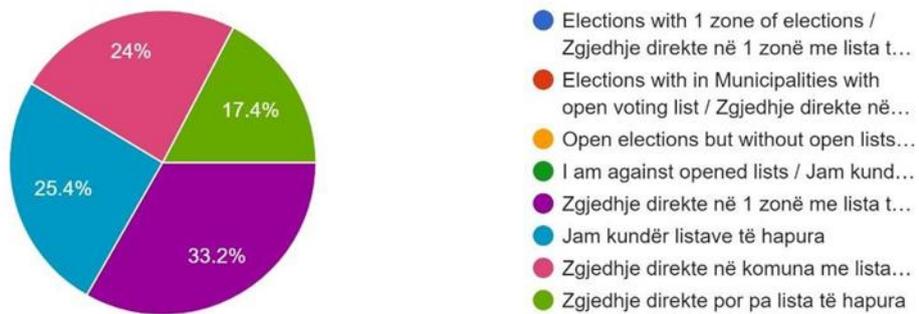


Figure 11: Enquiry for legislation one district voting preferences.

## 5. Conclusion

We have developed a method for recognizing the main impacting factors in elections by using Survey dataset that was originally collected within the study. The data are collected from Surveys because politicians and mainly voters, do not use substantially Twirrtter, facebook or Instagram and most of the analytics is based on input from Survey. In order to devise and asses the impacting factors we have devised most-important-problem algorithmic model to predict the outcome of 2021 Kosovo parliamentary elections and grounded it on the survey strategy.

An analysis of forecasting approach to elections and the performance metrics (variance) has been used.

The data has derived from an originally collected survey dataset that contains the impacting factors previously identified and assessed regarding the parliamentary elections in Kosovo February 2021 has been realized.

Discussion of results for each particular statistical analyses has been realized and the impacting fagcors will later be used to devise the prediction machine learning model.

## **References**

- [1] Czerlinski, J., Gigerenzer, G. & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97-118). Oxford University Press.
- [2] Gelman, A., Goel, S., Rivers, D., & Rothschild, D. (2016). The mythical swing voter. *Quarterly Journal of Political Science*, 11(1), 103–130.
- [3] Graefe, A. (2015). Improving forecasts using equally weighted predictors. *Journal of Business Research*, 68(8), 1792–1799.
- [4] Graefe, A., Armstrong, J. S., Jones, R. J. J., & Cuzán, A. G. (2014). Combining forecasts: An application to elections. *International Journal of Forecasting*, 30(1), 43–54.