Email: editor@ijerst.com or editor.ijerst@gmail.com

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

SENSITIVE ANALYSIS OF POLITICAL TWITTER USAGE IN VOTING FOR THE PUBLIC

[1] Ms Sadia Kausar, [2] Shaik Moazzam Mohiuddin, [3] Mohammed Mustafa, [4] Souban Mohiuddin
[1]Assistant Professor,Department of CSE-AIML, Lords Institute of Engineering & Technology.
[234] Student Department of CSE-AIML, Lords Institute of Engineering & Technology.

ABSTRACT

Classical political analysis tasks include surveying the public to gauge their opinions in order to make predictions about future elections (both national and local). But surveys may be costly to do, and if the sample isn't representative of the community, the findings might be skewed. We provide two novel approaches to using the findings of tweet sentiment analysis in survey research. Our first approach only relies on the sentiment analysis findings from tweets, which surpasses other widely-recognized methodologies. We also present a new hybrid method that estimates election outcomes using public opinion surveys and Twitter. The outcome of the 2023 Greek national election may be predicted using this approach, which allows for more precise, frequent, and cost-effective public opinion prediction. New opportunities for public opinion estimate using social media platforms were opened up by our technique, which showed less variance from the actual election outcomes than traditional public opinion surveys.

Terms related to this index include: political tweet analysis, popularity score, sentiment analysis, and Twitter data.

INTRODUCTION

Things that make up public opinion are assertions, thoughts, and ideas that are hard to put a number on. The level of support for specific political organizations, such parties and candidates, seems to be far easier to measure, however. For every i from 1 to n, there exists a mystery popularity score pi for each of the n (political) entities. Given the competitive nature of political voting, the political score, which is effectively the voting intention, shows what proportion of voters would choose entity (party or candidate) i above all other entities. It is originally unclear what the popularity ratings of each political entity are: pi, where i = 1,..., n. Several methods exist for estimating it, such as by analyzing data from social media platforms or by using traditional methods of population sampling and polling. The goal of any polling approach should be to estimate popularity scores pˆ = [pˆ1,..., pˆn]T that are as near as feasible to the unknown popularity numbers p = [p1,..., pn]T. These scores can only be known on rare occasions, such during an election or voting process. Assume P is the complete population set, and Pm and Po are subsets of P that represent (a) social media activists and (b) poll takers. Everyone in the set Pm writes political writings, and those texts often make reference to the n political entities in issue. The usage of hashtags on social media platforms allows users to link certain pieces of text to certain political figures or causes. Using text sentiment analysis, we may sort these messages into positive, neutral, and negative categories and assign numerical values to each category (ai, bi, and ci) for every political entity i = 1,..., n.
In this data analysis task, we need to find a way to predict ˆpm using the sentiment dataset S = n (ˆ ai, ˆ bi, ˆ ci), where i ranges from 1 to no. The corresponding estimates ˆp, ˆpm, ˆpo will vary as sets P, Pm, and Po change. We can estimate the popularity scores from S using ˆpo and historical measurements of ˆp, without conducting fresh conventional opinion polls, since voting outcomes are too rare compared to traditional polling results. With very small margins of error, traditional polling based on established Po has performed well over the years. Nevertheless, it has been an expensive effort to personally ask political questions and choose the sample set Po properly. So, if you want a quicker and less expensive way to estimate the distribution of popularity scores, social media sentiment analysis is a great option. Social media polling allows for more affordable and real-time distribution estimates of popularity scores as the number of individuals using social media to openly express their opinions on political issues continues to rise. In order to gauge popular sentiment on voting intentions, we offered two methods in this work for calculating ˆpm.
A lot of political rhetoric has come out of Twitter and other social media sites in the last ten years. Because of this, researchers felt compelled to look into the possibility of predicting election outcomes and public opinion based on data uploaded online

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

(tweets in the instance of Twitter). This approach gained significant traction during the 2016 US presidential elections and shown encouraging outcomes [1]. Two more two-party systems that performed similarly to this one were also subjected to the identical procedures [2, 3]. Method [2] was an attempt to refine the popularity measure suggested in [4] for the purpose of forecasting the outcome of the 2017 French presidential election's final round. It is not easy to extend these strategies to multi-party elections (n > 2). Controversial outcomes ensued from the various methods used to bridge the divide between two-party and multi-party elections. According to [5], the ratio of positive to negative remarks on a certain subject is called the sentiment score. It has been extensively used for predicting the outcomes of elections involving two or more political parties. For the UK general election of 2010, the real political terrain was mapped out in [6]. Using Twitter data alone to forecast political party popularity is clearly not going to cut it, according to this research. Methods similar to these have been used in [7]. Since most strategies for predicting general elections using Twitter have failed miserably, a hybrid prediction system using an election outcome regression model fed by several popularity score variables was developed [8]. Using the techniques outlined in [9], this system was trained using the results of traditional opinion polls. Data sentiment imbalance is a major problem in social media polling research. Political commentary on social media is problematic since it often comes from a small number of individuals who are likely prejudiced. It is now much more difficult to obtain accurate data for election prediction, as our study reveals that just 7% of the tweets collected are favorable. We thus suggest a new heuristic estimator for election results that relies heavily on unfavorable tweets. We also provide a new way to use historical election results and conventional polls to regress the distribution of popularity scores. Comparing our heuristic technique to others, we found that it deviated the least from the overall election outcomes. In addition, our hybrid approach beat out the state-of-the-art hybrid methods on literature and the final public opinion survey conducted by every major Greek polling firm before to the general election.

POLITICAL POPULARITY SCORE ESTIMATION BASED ON TWEET SENTIMENT ANALYSIS

Gauging popularity via heuristics

Using a population from either set Pm or Pm and Po, the remaining estimations in this study are based on the popularity scores ˆp, while maintaining generalizability. From political tweets that have been sentiment-labeled, we may heuristically estimate the popularity scores pi, where i = 1,..., n. First, we use tweet sentiment analysis [10] to automatically classify political tweets as good, neutral, or negative based on the party they belong to, as indicated by the hashtags. For each political entity (party) i = 1,..., n, let's say that there are a total of ai positive, neutral, and negative tweets represented by the n-dimensional vectors a, b, and c. Consider the total of all tweets, both positive and neutral, as vector d = a + b. In order to create the heuristic popularity score:

$$\hat{p_i}(\mathbf{c}, \mathbf{d}) = \frac{d_i}{d_\tau} \cdot (c_\tau - c_i). \tag{1}$$

The values of dτ and cτ are Pni=1 and ci, respectively. In essence, the sum of all negative tweets (excluding those from party i) is distributed according to the counts of positive and neutral comments for each party using the formula ˆ pi(c, d). Because the distribution of popularity scores should be Pn i=1 ˆ pi(c, d) = 1, we update this heuristic estimator appropriately and provide the Political Popularity Score Estimator (PPSE):

$$\hat{p_i}(\mathbf{c}, \mathbf{d}) = \frac{n \cdot d_i \cdot (c_\tau - c_i) + \mathbf{d}^T \mathbf{c}}{n \cdot c_\tau \cdot d_\tau} \tag{2}$$

Political popularity score regression from tweet sentiment analysis and past opinion polls

Opinion Poll Trends Regressor (OPTR)

The outcomes of using heuristics to estimate popularity scores have been encouraging over time. Unfortunately, as they were derived without considering ground truth or optimization criteria, their prediction accuracy is sub-optimal. So, they missed out on the benefits of Machine Learning's recent triumph. This being the case, we may compensate for the decline in precision by drawing on data from previously conducted traditional public opinion polls. In order to do this, we created a regression model that links changes in the outcomes of public opinion surveys to changes in the counts of positive, neutral, and negative variables: ajt, bjt, cjt, for j = 1,..., L, t = 1,..., Λ, during a certain time frame before two successive opinion polls. The total number of recorded conventional public opinion surveys is L, and index j denotes the chronological order of the poll. Our main premise is that if there is a major shift in the data measured on social media, the popularity ratings of the entities should also undergo a considerable change. The input to our model will be defined by subtracting the average positive, neutral,

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

and negative numbers from a certain time frame Λ prior to two successive opinion surveys.

$$\tilde{\mathbf{a}} = \frac{1}{\Lambda}\left(\sum_{t=1}^{\Lambda} a_{kt} - \sum_{t=1}^{\Lambda} a_{lt}\right), \quad (3)$$

$$\tilde{\mathbf{b}} = \frac{1}{\Lambda}\left(\sum_{t=1}^{\Lambda} b_{kt} - \sum_{t=1}^{\Lambda} b_{lt}\right), \quad (4)$$

$$\tilde{\mathbf{c}} = \frac{1}{\Lambda}\left(\sum_{t=1}^{\Lambda} c_{kt} - \sum_{t=1}^{\Lambda} c_{lt}\right). \quad (5)$$

This mapping is executed and the change in popularity scores are estimated by feeding it into a basic regression model.

$$\hat{\mathbf{r}} = f(\mathbf{x}; \mathbf{w}), \quad (6)$$

where the input vector x = h ˜a||˜ b||˜c i is of dimension 3n and f is the model regression function, using w as its parameters. Two successive conventional polls are indicated by variables k and l. The MSE loss is used throughout the training method. After training on sample data D = {ˆpok − ˆpol, xkl}, k, l = 1,..., L, the regression model (6) may take the previous measurement and add ˆr to get the next popularity score estimate ˆp.

$$\hat{\mathbf{p}}_{t+1} = \hat{\mathbf{p}}_t + \hat{\mathbf{r}}_t, \quad (7)$$

The time spot being computed, often in days, is denoted by t. Assuming that the results of the traditional polls would be affected by the inverse of the change in the social media statistics, variable k might be either k = l + 1 or k = l − 1 for data augmentation purposes. We choose to use the difference between two successive opinion polls ˆpok − ˆpol, where k = l ± 1, to counteract the bias that is created by traditional public opinion surveys, rather than relying only on their findings ˆpok. OPTR is very much like the approach suggested in [8]. In contrast to our approach, the second one uses real opinion poll estimates as its output and uses characteristics generated by heuristic estimators as its regression input. Consequently, their findings don't account for the inherent bias in opinion surveys and don't examine the political system's constituents (parties) as interdependent entities. As shown in(7), the suggested strategy involves basing fresh predictions on the prior ones. Due to the inherent bias in traditional public opinion surveys, it is necessary to derive starting values from real election results. Grouping opinion polls for the regression model (2.2.2) There will always be differences in the results of different businesses' traditional public opinion surveys, and these differences might affect the accuracy of the aforementioned political popularity score estimate. The selected method involves classifying similar polls taken during the same time period into distinct categories based on the degree to which they deviate from the estimates provided by other survey firms; these categories will then be used in the regression model (6). Based on a survey performed by firm z = 1,...,m on date tk, let's say that uzi(tk) is the estimated popularity score of party i. Because different polling organizations use different dates for their surveys, we use the formula to perform linear interpolation for each political entity between two successive surveys done by the same business for a given date t, where tk < t ≤ tk+1.

$$u_{zi}(t) = u_{zi}(t_k) + (t - t_k)\frac{u_{zi}(t_{k+1}) - u_{zi}(t_k)}{t_{k+1} - t_k}. \quad (8)$$

After that, on date t, we may add up the Mean Absolute Errors (MAE) eŎ(t) from the polls conducted by firm Ŏ and the other polling companies:

$$e_\zeta(t) = \sum_{z=1, z\neq\zeta}^{m} \frac{\sum_{i=1}^{n} |u_{\zeta i}(t) - u_{zi}(t)|}{n}, \quad (9)$$

such that z ranges from 1 to m. Another issue that is affecting our regressor is the fact that various firms have varying public opinion estimates for the same time period. This is why we used the weighted average based on the polls' individual mistakes to combine public opinion surveys that were conducted fewer than d days apart. The opinion polling problem requirements determine the hyperparameter d.

EVALUATION POLITICAL POPULARITY SCORE ESTIMATION METHODS

Data Gathering

Using the Twitter API, we were able to collect 1,001,836 tweets about six Greek political legislative parties between June 25, 2022, and June 25, 2023. Using the Transformer technique, which was suggested in [11], all tweets have been categorized as neutral, positive, or negative. This method has a 79% accuracy rate in sentiment classification when evaluated on ground truth Greek political tweets [12]. We trained our regression model (6) and validated our suggested procedures using 35 public opinion surveys that we collected throughout the data collection period. The general elections in Greece were conducted on May 21, 2023, and June 25, 2023,

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

as per the Greek constitution. 3.2. Evaluation of heuristics for estimating political popularity scores We set out to test our PPSE estimator using political Greek Twitter data and compare it to five other heuristic estimators and [2, 4, 5, 8, 13] that have been either suggested as estimators or used as features in regression models. Therefore, we determined the popularity score for each Greek parliamentary party during the data collecting period based on the aforementioned estimators. Next, we computed the Mean Absolute Error (MAE) to evaluate the various heuristic estimators. This measure is derived from the average deviation of each estimator and the ground truth, which in this case is the results of the general election. Since the sum of certain estimators is not 1 for all n entities, we began by normalizing them: ´ pi = P pˆi n i=1 ˆ pi .

Table 1. MAE between heuristic estimators' results and the results of the Greek general elections of May(21/5/2023) and June (25/6/2023).

Estimators	300 days		200 days		100 days	
	May	June	May	June	May	June
[5]	21.2%	20.71%	21.23%	20.71%	20.94%	20.1%
[8]	19.17%	18.66%	19.18%	18.38%	18.88%	18.37%
[13]	9.2%	9.94%	9.2%	10.97%	10.47%	11.33%
[4]	10.43%	11.27%	10.43%	11.81%	11.55%	11.98%
[2]	9.35%	8.79%	9.35%	8.2%	9.17%	**7.75%**
PPSE (proposed)	**7.06%**	**7.31%**	**7.15%**	**7.62%**	**7.12%**	7.76%

Table 1 presents the testing results during 3 periods, starting from 21 May 2023 and 25 June 2023 for a time window

Table 2. MAE between estimators (OPTR and method [8]), the last recorded opinion poll from different polling companies and the Greek general elections results 25/6/2023.

Methods/Polling Companies	MAE
METRON ANALYSIS	2.17%
MRB	1.89%
MARC	1.63%
GPO	1.57%
PULSE	1.54%
Method [8]	1.42%
OPTR	**1.09%**

between one hundred and three hundred days in the past, for the two separate election dates. When compared to competing estimators, ours performs better on the majority of the examined windows. It should be mentioned that our heuristic estimator was the only one to accurately anticipate the actual party ranking (ND > SYRIZA > KINAL > KKE > ELLINIKI LISI > MERA25) based on the election records. Still, no estimate has been able to reliably forecast the true vote shares. Because surveys are able to choose a more representative cross-section of society than Twitter, this might happen. Therefore, the suggested OPTR technique, which offers better prediction of election results (as will be discussed later), is the best option practically speaking. Evaluation of the OPTR model (3.3) Neither the poll results from the last election nor the tweets from that time period were accessible because data gathering began in June 2022. The good news is that we were able to put our strategy to the test since there were two general elections. Our strategy essentially tracked the changes in popularity scores from the first election on May 21, 2023, to the second election on June 25, 2023, using the baseline numbers from that date. We contrasted our procedure with that of other polling firms and the outcomes of other methods [8].

As shown in Table 2, the actual election results from 25/6/2023 are compared with the latest recorded opinion poll of each firm and our suggested approach, the [8] method. By comparing it to the method suggested in [8] and every traditional opinion poll conducted in the two weeks leading up to the election, opinion poll regressor (OPTR) clearly comes out on top. OPTR was trained using skewed, noisy opinion surveys that were released before to the first election day (21/5/2023). Despite the fact that our method gleaned political trends from those unreliable samples, when combined with the results of the first election, it outperformed all other methods and opinion polls—all of which did not include surveys released after May 21, 2023. Based on the calculations made from equation (9), Figure 1 displays the mistake that each firm has incurred up to the first election date. This inaccuracy is in line with Table 2, providing further evidence that the variance across polling firms may be used to filter out irrelevant differences in training data.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

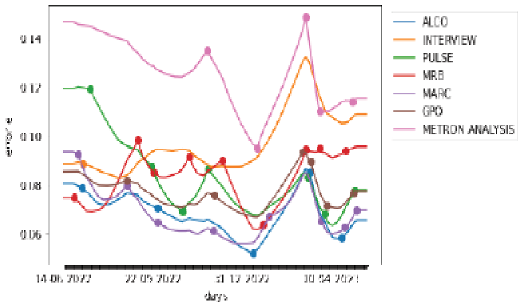

Fig. 1. Poll error (e), from other public opinion polls, throughout the number of days since the 14^{th} of June. Dots indicate the last day when each poll was conducted.

CONCLUSION

In this study, we provide a heuristic and a regression approach for calculating political popularity ratings using sentiment analysis in social media data. Both of these provide quite accurate estimates of how popular politicians are. The regression-based approach outperforms the heuristic one, but it also requires familiarity with historical election outcomes and opinion poll data. There is still a significant gap between heuristic popularity estimators and hybrid ones that use social media and previous conventional polls. However, with the advancement of Natural Language Processing (NLP) tools, the accuracy of social media political forecasting should increase. In the meanwhile, hybrid regression techniques are more effective. In contrast to more traditional opinion polling services, our findings showed that a hybrid approach combining Twitter data with surveys produced superior results. Using social media as a main data source, this article suggests a new way to analyze politics. Our research shows that this method is a huge step forward in the sector, as it beats out more conventional methods of gathering public opinion.

REFERENCES

[1] Brandon Joyce and Jing Deng, "Sentiment analysis of tweets for the 2016 us presidential election," in 2017 IEEE MIT Undergraduate Research Technology Conference (URTC), 2017, pp. 1–4.

[2] Lei Wang and John Q. Gan, "Prediction of the 2017 french election based on twitter data analysis," in 2017 9th Computer Science and Electronic Engineering (CEEC), 2017, pp. 89–93.

[3] Alexandre Bovet, Flaviano Morone, and Hernan A. Makse, "Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump," Scientific Reports, vol. 8, no. 1, pp. 8673, June 2018.

[4] Andranik Tumasjan, Timm Oliver Sprenger, Philipp G. Sandner, and Isabell M. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment," Proceedings of the International AAAI Conference on Web and Social Media, 2010.

[5] Brendan O'Connor, Ramnath Balasubramanyan, Bryan Routledge, and Noah Smith, "From tweets to polls: Linking text sentiment to public opinion time series," International AAAI Conference on Weblogs and Social Media, vol. 11, 01 2010.

[6] Yulan He, Hassan Saif, Zhongyu Wei, and Kam-Fai Wong, "Quantising opinions for political tweets analysis," in LREC, 2012.

[7] Pete Burnap, Rachel Gibson, Luke Sloan, Rosalynd Southern, and Matthew Williams, "140 characters to victory?: Using twitter to predict the uk 2015 general election," Electoral Studies, vol. 41, pp. 230–233, 2016.

[8] Adam Bermingham and Alan Smeaton, "On using Twitter to monitor political sentiment and predict election results," in Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011), Chiang Mai, Thailand, Nov. 2011, pp. 2–10.

[9] Sitaram Asur and Bernardo A. Huberman, "Predicting the future with social media," in 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2010, vol. 1, pp. 492–499.

[10] Dionisis Karamouzas, Ioannis Mademlis, and Ioannis Pitas, "Public opinion monitoring through collective semantic analysis of tweets," Social Network Analysis and Mining, 07 2022.

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, L ukasz Kaiser, and Illia Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. 2017, vol. 30, Curran Associates, Inc.

[12] Emmanouil Patsiouras, Ioanna Koroni, Ioannis Mademlis, and Ioannis Pitas, "Greekpolitics: Sentiment analysis on greek politically charged tweets," 31st European Signal Processing Conference (EUSIPCO), 09 2023.

[13] Barkha Bansal and Sangeet Srivastava, "On predicting elections with hybrid topic based sentiment analysis of tweets," Procedia Computer Science, vol. 135, pp. 346– 353, 2018.