
ISSN: 2321-2152
IJMECE
International Journal of modern electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

Public opinion polling using political sentiment analysis on Twitter

[1]Mr. P. S. S. K. Sarma,[2] Charan Chintalapati,

[1]Associate Professor, Department of MCA, Rajamahendri Institute of Engineering & Technology. Bhoopalapatnam, Near Pidimgoyyi, Rajahmundry,E.G.Dist.A.P. 533107.

[2]Student, Department of MCA, Rajamahendri Institute of Engineering & Technology. Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

Abstract:

Predicting the results of local and state elections by surveying the general population is an age-old practice in political science. However, polls might produce biased results due to inaccuracies in population sampling and can be expensive to run. This article presents two unique ways to poll the general public using the results of sentiment analysis on tweets. When compared to a public opinion poll with many knowledgeable participants, our first method—which is based only on tweet sentiment analysis—performs better. Every member of the set Pm often makes reference to the political entities 1,..., n in their political writings. Hashtags are a fantastic tool for tagging certain social media posts with political people or subjects. Text sentiment analysis categorizes messages into positive, neutral, and negative sentiment classes. For each political entity i = 1,..., n, the values ai, bi, and ci may be found using this method. With i ranging from 1 to n, the present data analysis task is to get the sentiment score from the dataset S = (aˆ, bˆ, cˆ). Using survey and social media data, we can make educated guesses about the results of the next election. Using this method, we can gauge public sentiment more often, accurately, and inexpensively, allowing us to forecast the 2023 Greek national election result. By demonstrating that our method differed less from the real election results than conventional polls, we paved the way for new ways of gauging public sentiment using social media. Analytics on sentiment, popularity, election forecasting, and political tweets are all part of the index.

INTRODUCTION

Assertions, ideas, and concepts that are difficult to quantify comprise what is known as popular opinion. It seems to be far simpler to measure the popularity of certain political entities, such as political parties and individuals. There is an unknown popularity score pi for each of the n political entities for every integer i from 1 to n. Since political voting is inherently competitive, the political score—which is almost identical to the voting intention—reveals what percentage of the public would prioritize entity (party or candidate) i above all other entities. Pi, where i = 1,..., n, represents the popularity ratings of each political entity. At first, these numbers are ambiguous. It might be approximated in a number of ways; two examples are the use of social media data analysis and the use of traditional techniques of population sampling and polling. On rare circumstances, such during an election or voting process, the unknown popularity scores, represented as p = [p1,..., pn]T, might be known. Any polling technique's estimate of popularity scores, pˆ = [pˆ1,..., pˆn]T, should roughly approximate these scores. Whether P represents the whole population or a subgroup that is politically active on social media (Pm) or not (Po) determines the variation in the estimates pˆ, pˆm, and pˆo. Since vote outcomes are uncommon and traditional polling results are more prevalent, we may estimate the popularity scores from S using pˆo and previous measures of pˆ instead of conducting new conventional opinion surveys. In the past, satisfactory findings with modest margins of error have been generated using traditional polling using set Po. It has been a costly endeavor to choose the appropriate sample set Po and ask political questions by hand. Social media sentiment analysis

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

might therefore provide an alternative, more efficient, and cheaper method of determining the dispersion of popularity ratings (pˆm). As more and more people use social media to publicly voice their opinions on political matters, social media polling provides an opportunity for more economical and real-time popularity score dispersion estimations. This paper presents two ways to compute pˆm, which is a measure of public opinion based on voting intention.

Twitter and other social media have been heavily used by political campaigns for communication throughout the last decade. As a result, scientists felt forced to investigate if it was possible to gauge public sentiment and forecast election results using just data collected from the internet (in this case, tweets). There were hopeful results from this trend, which gained considerable momentum during the 2016 US presidential elections [1]. Similar two-party arrangements with two members each also fared quite well [2, 3]. The goal of Method [2] was to improve upon the popularity indicator proposed in [4] so that it might predict the results of the last round of the 2017 French presidential election. However, modifying these approaches for multi-party elections is hardly a walk in the park. Attempts to combine features of two-party and multi-party systems have produced contentious results. An approach to calculating the emotion score might include dividing the proportion of positive

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

also included critical comments on a topic in [5]. Using this method, you can get neutral comment counts. Because the popularity score distribution is often used to forecast the results of two- or multi-party elections, we modify this heuristic prediction to ensure it follows the assumption that Γnpˆi(c, d) = 1. The actual political climate for the 2010 UK general election was drawn out by researchers there [6]. According to this study, it is evident that using Twitter data alone to predict party support is not going to be effective. To compare with existing methods, we provide the Political Popularity Score Estimator (PPSE), which makes proper use of an estimator:

2. The Fascination of Politics TWITTER SENTIMENT ANALYSIS FOR SCORE ESTIMATION

Over time, the heuristic popularity score estimates made by the OpinionPollTrendsRegressor (OPTR) have demonstrated promising results. The inferior prediction accuracy is a result of their failure to include ground truth or optimization criteria into their derivation. Because of this, they were unable to take advantage of Machine Learning's current prosperity. Given this, we may use results from older, more conventional public opinion surveys to make up for the loss of accuracy. Our methodology included developing a regression model that establishes a connection between shifts in public opinion survey results and changes in the counts of positive, neutral, and negative sentiment (ajt, bjt, and cjt, respectively) during a certain time span before two subsequent opinion polls. Index j indicates the chronological order of the poll, and L is the total number of recorded traditional public opinion polls. We based our method on the fundamental assumption that entities' popularity ratings should vary drastically when there are big changes in the social media data that is tracked. The input of our model may now be determined by subtracting the average positive, neutral, and negative numbers of a specified time period Λ from two consecutive opinion surveys.

The popularity ratings pˇ, used for all future predictions in this work, may be derived from either the set Pm or a mix of Pm and Po, ensuring that generalizability is not compromised. These are the ratings for ease of use.

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

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

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

The regression model may predict the next popularity score p by adding ˆr to the previous measurement: ˆ pt+1=ˆ pt+ˆrt, where t is the exact time point being computed (typically days) and D is the sample data. For data augmentation purposes, variable k may be either k=l+1 or k=l−1, supposing that the inverse change in social media statistics will provide the exact opposite change in traditional poll findings. Instead of utilizing the findings of traditional public opinion surveys directly, we will use the difference between two successive polls, k=l±1, to counteract the bias that these polls create. The approach suggested in[8] is comparable to OPTR. In contrast to our technique, the latter uses real opinion poll estimates in its output and relies on characteristics derived by heuristic estimators for its regression input. Consequently, their findings do not account for the fact that political parties are dependent organizations and do not account for the fact that polls introduce bias. New estimates will be provided in relation to the prior ones, as shown in (7), for the suggested technique. Because of the inherent bias in traditional public opinion surveys, this necessitates the use of starting values derived from real election outcomes. 2.2.2. Grouping opinions for the regression model The aforementioned method of estimating political popularity might be vulnerable to the inevitable differences between different conventional public opinion surveys carried out by different organizations. The recommended methods for combining many polls taken within the same time period into a single regression model based on how far each poll deviated from the average (6). Assume for the sake of argument that uzi(tk) is the prediction of the popularity score of party i in a poll carried out by business z=1,...,m at date tk. We use linear interpolation for each political leaning between two subsequent polls performed by the same business for a given date, where tk≤tk+1, since the poll dates change between polling organizations. The formula is:

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

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

Our regressor is also experiencing issues due to estimates made by other firms over the same time. In order to achieve this goal, the weighted average was calculated according to the corresponding mistakes when two or more public opinion surveys were conducted fewer than ten days apart. It is the requirements of the opinion polling problem that determine the hyperparameter. 3.

EVALUATIONPOLITICALPOPULARIT YSCORE ESTIMATIONMETHODS

3.1. Collecting Data From June 25, 2022, to June 25, 2023, a total of 1,001,836 tweets were collected about six Greek political legislative parties using the Twitter API. The Transformer approach, which was evaluated on ground truth Greek political tweets [12] and shown to demonstrate 79% sentiment detection accuracy, has been proposed in [11] and used to identify all tweets as neutral, positive, or negative. We trained our regression model (6) and validated our suggested procedures using 35 public opinion surveys that we collected throughout the data collection period. According to the rules of the Greek constitution, the Greek general elections were conducted on 21/5/2023 and 25/6/2023. 3.2. Heuristic political popularity score timators compared We set out to assess and contrast our estimator (PPSE) using political Greek Twitter data with five other fuzzy estimators and [2,4,5,8,13] that have previously been suggested as estimators or used as features in regression models. During the data collecting period, we calculated the popularity score for every Greek parliamentary party according to the estimators described earlier. We then computed the Mean Absolute Error to compare the various heuris estimator outputs. After that,

(MAE), with the term "ground truth" meaning the average divergence of the results from both the polls and the general election. We first normalized them: 'pi= ˆ pi n i=1 ˆ pi,' since certain estimators don't sum to 1 for all entities. The results of the Greek general elections in May (2023) and June (25/6/2023) were compared with the MAE between the heuristic estimators' findings in Table 1.

June (25/6/2023).

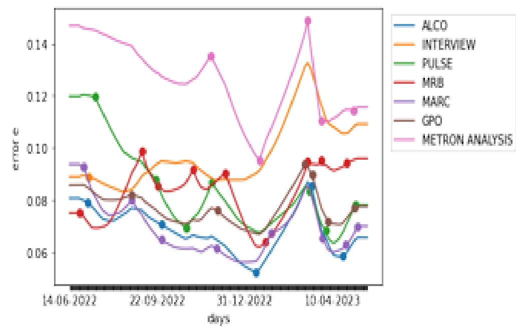
Estimators	300 days May	300 days June	200 days May	200 days June	100 days May	100 days June
[5]	21.2%	20.71%	21.23%	20.71%	20.94%	20.1%
[8]	19.17%	18.66%	19.18%	18.38%	18.88%	18.37%
[13]	9.2%	9.94%	9.2%	10.97%	10.47%	11.33%
[4]	10.43%	11.27%	10.43%	11.81%	11.55%	11.98%
[2]	9.35%	8.79%	9.35%	8.2%	9.17%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

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

between one hundred and three hundred days in the past, for the two separate election

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

dates. When compared to competing estimators, ours performs better on the majority of the examined windows. It is worth mentioning that our heuristic estimator was the only one to accurately anticipate the actual party ranking (ND > SYRIZA > KINAL > KKE > ELLINIKILISI > MERA25) according to the election results. But predicting the real vote shares is a challenge for all the estimators. This might happen because surveys are better able to choose a representative cross-section of society than Twitter. As will be shown in the section that follows, the suggested OPTRmethod offers better prediction of election results, hence it is clearly the preferred option. 3.3. Assessment of the OPTR model Both the poll results and the tweets from the last election were inaccessible because data gathering began in June 2022. Fortunately, we were able to test our strategy since there were two general elections. From the first set of election results on May 21, 2023, to the second set of results on June 25, 2023, our algorithm determined the changes in popularity scores. Various polling firms and their methods' outcomes are contrasted with ours [8]. Based on the actual election results from 25/6/2023, Table 2 shows the estimates for our suggested technique, the [8] method, and the latest recorded opinion poll of each firm before the election date MAE. It is clear that OPTR performs better than both the method suggested in [8] and all of the traditional opinion surveys conducted in the two weeks leading up to the election. The findings showed that the noisy opinion surveys used to train OPTR were biased since they were collected before the first election date (21/5/2023). Without using any polls released after 21/5/2023, our method outperforms all others when combined with the first election's results and the noisy samples from which our methodology learnt the political trends. The inaccuracy of each corporation relative to the others up to the first election date, as determined from (9), is shown in Figure 1. This mistake is in agreement with Table 2, providing further evidence that the variance across polling firms may be used to remove undesired fluctuations from training data.



As the number of days after June 14th increases, the poll error (e) from other public opinion surveys decreases (Fig. 1). The last day that each poll was conducted is shown by the dots. 4. FINAL THOUGHTS This study presents two novel methods—a heuristic and a regression—for calculating political popularity ratings from sentiment analysis of social media data. They are both useful tools for gauging public support for various political candidates. Although it requires familiarity with historical election outcomes and opinion poll data, the regression-based approach outperforms the heuristic one in terms of accuracy. The results we get from political forecasting through social media should become more and more accurate as Natural Language Processing (NLP) tools advance. However, for the time being, hybrid regression techniques outperform heuristic popularity estimators in terms of accuracy. Our trials, however, showed that a hybrid approach combining Twitter data with opinion surveys outperformed more traditional methods used by polling firms. 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 discipline, as it beats out more conventional public opinion surveys.

REFERENCES

[1]. [1] Brandon Joyce and Jing Deng, "Sentiment analysis of tweets for the 2016 us presidential election," in 2017

IEEEMITUndergraduateResearchTechnologyConfer ence (URTC), 2017, pp. 1–4.

[2]. [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]. [3] Alexandre Bovet, Flaviano Morone, and Hernan A. Makse, "Validation of Twitter opinion trends with na tional polling aggregates: Hillary Clinton vs Donald Trump," Scientific Reports, vol. 8, no. 1, pp. 8673, June 2018

[4]. . [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 Con ference on Web and Social Media, 2010.

[5]. [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]. [6] Yulan He, Hassan Saif, Zhongyu Wei, and Kam-Fai Wong, "Quantising opinions for political tweets anal ysis," in LREC, 2012.

[7]. [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]. [8] AdamBerminghamandAlanSmeaton, "OnusingTwit ter 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]. [9] Sitaram Asur and Bernardo A. Huberman, "Predicting the future with social media," in 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intel ligent Agent Technology, 2010, vol. 1, pp. 492–499.

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

[11]. [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]. [12] Emmanouil Patsiouras, Ioanna Koroni, Ioannis Madem lis, and Ioannis Pitas, "Greekpolitics: Sentiment analy sis on greek politically charged tweets," 31st European Signal Processing Conference (EUSIPCO), 09 2023.

[13]. [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.