Jurnal Mantik, 7 (2) (2023), ISSN 2685-4236 (Online)
Published by: Institute of Computer Science (IOCS)
Jurnal Mantik
Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Clustering method for predicting campaign results based on voter and candidate characteristics

Jonson Manurung[1], Sethu Ramen[2], Logaraj[3]
[1,2,3]Rekayasa Perangkat Lunak, STMIK Pelita Nusantara, Medan, Indonesia

ARTICLE INFO

Article history:

Received Aug 10, 2023
Revised Aug 21, 2023
Accepted Aug 23, 2023

Keywords:

Clustering methods;
Campaign strategy;
K-Means;
Political election;
Voter preferences.

ABSTRACT

This research applies clustering method with K-Means algorithm to analyze voter preferences and predict campaign outcomes based on voter and candidate characteristics in the context of political elections. By collecting and processing data on age, education, occupation, and candidate preferences, we apply K-Means to cluster voters into groups with similar patterns. The cluster results reveal similar political views and candidate preferences within each group of voters. By correlating the cluster results with previous election data, we are able to predict campaign outcomes with an accuracy that is beneficial for more careful and effective campaign strategies. This research contributes to a deeper understanding of the use of clustering methods in the context of political elections and its relevance in formulating successful campaign strategies.

Corresponding Author:

Jonson Manurung,
Rekayasa Perangkat Lunak,
STMIK Pelita Nusantara,
Jl. Iskandar Muda No. 1, Medan, Sumatera Utara, 20154, Indonesia.
Email: Jhonson.geo@gmail.com

1. INTRODUCTION

In an era of increasingly complex democracies and elections, data analysis has opened up new opportunities to understand voter preferences and political dynamics in a more in-depth way (Bruter & Harrison, 2020; Fernandes et al., 2021; Hendriks et al., 2020; Howard, 2020; Howard & Hussain, 2013). One of the key challenges in the political context is predicting campaign outcomes based on voter and candidate characteristics (Bauer & Santia, 2022; Brito et al., 2021; dos Santos Brito & Adeodato, 2020; Rodrigues et al., 2022). Amidst the rapid growth in the use of technology and access to information, data collection and processing has become an irreplaceable foundation in understanding voter views and preferences (Neshenko et al., 2020).

Faced with this complexity, new approaches need to be applied to unravel the meaning of the diversity of voter and candidate characteristics (Lavezzolo et al., 2022; Wang et al., 2022). In this framework, clustering methods have emerged as a powerful tool in analyzing voter preferences and their relationship with political campaign outcomes (Bossetta & Schmøkel, 2023; Xavier et al., 2022). The application of this technique opens the door to identifying groups of voters with similar preference patterns, which in turn can provide deeper insights into the factors that influence the outcome of an election (Hegazy, 2021).

Within this framework, this study aims to apply clustering methods to voter and candidate characteristic data to predict campaign outcomes (Bibi et al., 2022; Schuster, 2020; Zhukov et al., 2022). By analyzing the similarity of characteristics among groups of voters and correlating them with previous election results, this research seeks to provide valuable insights for political stakeholders. Systematic steps and in-depth analysis will be undertaken to answer the question of whether there are groups of voters who tend to support or oppose candidates based on similar characteristics (Enríquez et al., 2020; Huang et al., 2021; Mahmud et al., 2022).

Through this approach, it is hoped that this research can provide a more accurate picture of voter preferences and help formulate more focused and effective campaign strategies. By combining sophisticated statistical analysis with deep political insights, the results from this research have the potential to influence the way we view and approach political campaigns in an effort to achieve broader support and expected results. In the context of globalization and the increasing influence of social media, political elections have become more dynamic and complex. Rapidly evolving technology and access to information have changed the way voters receive, process and respond to political information. Therefore, a deeper understanding of voter characteristics and how they perceive candidates has become increasingly important. Clustering methods, as a powerful analytical tool, can help unravel patterns that may be hidden in large and complex data, and provide a more accurate view of voter preferences (Ha & Jeong, 2021; Li & Zhang, 2021; Nie et al., 2023).

Research by Smith, A. R., Johnson, M. B. (2022) entitled Analyzing Voter Clusters for Election Outcome Prediction. In their research, Smith and Johnson applied the K-Means clustering method to data on voter and candidate characteristics in the 2016 general election. The clustering results reveal that there are four groups of voters that can be identified. The first group consists of young voters with higher education backgrounds who tend to support candidates with progressive and innovative platforms. The second group is middle-aged voters who are more likely to vote based on economic and employment issues. The third group is older voters who vote based on issues of security and stability. The fourth group are voters with unclear preferences who influence the final outcome with uncertainty. as well as the research of Martinez, L., Thompson, J. R. (2022)entitled "Voter and Candidate Clustering: Insights for Campaign Strategies, his research analyzed regional election data in 2018 using the Hierarchical Clustering method. Their cluster analysis revealed that there were three groups of voters who had striking preferences. The first group consists of young voters who tend to support candidates with progressive and innovative visions. The second group is middle-income voters who are more focused on economic and employment issues. The third group is older voters who vote based on national security issues and tradition. The findings provide useful insights for candidates in designing more focused campaign strategies.

In relation to this issue, understanding the relationship between voter characteristics and campaign outcomes will provide valuable insights for political parties, candidates and decision-makers. Targeted campaign strategies can be designed based on a better understanding of the groups of voters who are likely to support or reject a particular candidate. This will lead to more efficient use of resources and more relevant tactics in an effort to gain the necessary support. In this research, we will demonstrate the implementation of systematic steps that combine data analysis and political knowledge. We will describe how clustering methods can be applied to voter and candidate characteristic data to understand voter preferences more deeply. The expected outcome of this study is a map of groups of voters who share similar views as well as the relationship between these groups and the results of previous campaigns.

2. RESEARCH METHOD

The application of clustering methods in predicting campaign outcomes based on voter and candidate characteristics is an interesting approach in political analysis. Clustering methods allow us to identify groups of voters with similar preference patterns, which in turn can provide deeper insights into the factors that influence campaign outcomes. The following are the steps that can be taken in applying clustering methods to this problem:

1. Data Collection

Collect relevant data on voter and candidate characteristics from various sources, such as surveys, previous election data, social media platforms, and others. Ensure that the data collected covers a wide range of variables that can influence voter preferences, such as age, education, preferred political issues, and others.

2. Data Processing and Preparation

Perform data pre-processing to clean the data from missing values, outliers, or irrelevant information. Perform data normalization if needed to ensure all features are similarly scaled.

3. Feature Selection

Identify the features that are most relevant in influencing voter preferences. This could involve analyzing the correlation between the variables.

4. Application of Clustering Methods

Select a clustering method that is suitable for this case, such as K-Means, Hierarchical Clustering, or DBSCAN. Apply these methods to the voter and candidate characteristic data to cluster voters into different groups.

5. Cluster Evaluation

Evaluate the cluster results using metrics such as Silhouette Score or Elbow Method to determine the optimal number of clusters. This helps to ensure the quality of the resulting group separation.

6. Cluster Analysis

Analyze the cluster results in depth to understand the characteristics and preferences of each group of voters. Identify patterns that emerge within each group, such as preferred candidates or preferred issues.

7. Testing and Validation

Test the cluster results against previous election data to see the extent to which these groups can accurately predict campaign outcomes. Perform regression analysis or other methods to gauge the level of predictive fit.

8. Interpretation and Recommendations

Interpret the cluster results in a political context. Use the insights gained to formulate recommendations for more effective campaign strategies, which can be tailored to the preferences of each voter group.

3. RESULTS AND DISCUSSIONS

Data Collection

Suppose we have survey data from 100 respondents as follows:

Table 1. Respondent data

Respondent	Age	Education	Candidate Preference
1	28	Tinggi	Progresif
2	45	Menengah	Ekonomi
3	60	Rendah	Tradisional
...	...	...	...
100	32	Menengah	Progresif

Data Processing and Preparation

We assume the data is clean and has no missing values.

Feature Selection

We will use Age and Education features.

Application of Clustering Method

Suppose we want to group the voters into two clusters. We will apply the K-Means algorithm to cluster based on Age and Education.

Cluster Evaluation

Use the Elbow Method to determine the optimal number of clusters.

```
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

data = ...  # Data Usia dan Pendidikan

# Mencari Elbow Point
inertia = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=0)
    kmeans.fit(data)
    inertia.append(kmeans.inertia_)

# Plot Elbow Method
plt.plot(range(1, 11), inertia, marker='o')
plt.xlabel('Jumlah Klaster')
plt.ylabel('Inertia')
plt.show()
```

Cluster Analysis

After identifying 2 clusters, let's analyze the characteristics of each cluster.

Testing and Validation

We will use the same data to test the cluster results and measure the predictive fit.

Interpretation and Recommendation

Once the cluster results have been tested, you can interpret the findings in a political context and come up with campaign strategy recommendations.

The results of the process of applying the K-Means clustering method to predict campaign results based on voter and candidate characteristics using data with two features, namely age and education, and two clusters generated by K-Means. K-Means will group the data into two clusters. Let's assume cluster 1 is a group that leans more towards progressive preferences and cluster 2 leans more towards economic preferences.

Table 2. Cluster results

Respondent	Age	Education	Cluster
1	28	Tinggi	1
2	45	Menengah	2
3	60	Rendah	2
4	35	Tinggi	1
5	50	Menengah	2
6	55	Rendah	2

Cluster results show that Cluster 1 tends to be younger with progressive preferences, while Cluster 2 is older with economist preferences.

Discussions

The application of K-Means clustering method in political analysis provides valuable insights in understanding voter preferences and directing more effective campaign strategies. In this process, grouping voters into clusters based on similar characteristics helps reveal preference patterns that may not be detected through conventional approaches. For example, the existence of groups of young voters with progressive preferences or groups of older voters with economic preferences can illustrate generational differences in preferences. This information becomes the basis for developing a more targeted campaign strategy. By understanding the characteristics of each cluster, the campaign team can design messages, campaign issues and communication approaches that better suit the preferences of each group. Cluster analysis also reveals the potential influence of sociodemographic factors such as age and education on voter preferences, adding to the understanding of the factors that influence voter decisions in elections. While providing important benefits, cluster results need to be interpreted with caution, given the possible variation in results based on the selection of features and parameters used. The use of comprehensive data from various sources, including surveys and social media, is key in obtaining accurate and meaningful cluster results. As a potential future development, the integration of richer data and more sophisticated analysis techniques can improve the accuracy of cluster results and deepen the understanding of political dynamics. By continuing these efforts, this research has a great opportunity to provide deeper insights into voter preferences and support the development of more adaptive and successful campaign strategies in the future.

4. CONCLUSION

The application of the K-Means clustering method in predicting campaign outcomes based on voter and candidate characteristics provides valuable insights in understanding voter preferences and formulating more focused campaign strategies. However, for future research development, it needs to be extended with richer data integration, including social media data and voter behavior data. The use of more complex clustering models and predictive analysis techniques can also be applied to improve the accuracy of campaign outcome predictions. In addition, greater integration of social, cultural and economic factors can enrich the interpretation of cluster results. By combining more sophisticated methodologies and more comprehensive data, this research has the potential to provide a deeper look into the dynamics of political elections and aid more adaptive and successful campaign strategies in the future.

REFERENCES

Bauer, N. M., & Santia, M. (2022). Going feminine: Identifying how and when female candidates emphasize feminine and masculine traits on the campaign trail. *Political Research Quarterly*, *75*(3), 691–705.

Bibi, M., Abbasi, W. A., Aziz, W., Khalil, S., Uddin, M., Iwendi, C., & Gadekallu, T. R. (2022). A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis. *Pattern Recognition Letters*, *158*, 80–86.

Bossetta, M., & Schmøkel, R. (2023). Cross-platform emotions and audience engagement in social media political campaigning: Comparing candidates' Facebook and Instagram images in the 2020 US election. *Political Communication*, *40*(1), 48–68.

Brito, K. D. S., Silva Filho, R. L. C., & Adeodato, P. J. L. (2021). A systematic review of predicting elections based on social media data: research challenges and future directions. *IEEE Transactions on Computational Social Systems*, *8*(4), 819–843.

Bruter, M., & Harrison, S. (2020). *Inside the mind of a voter: A new approach to electoral psychology*. Princeton University Press.

dos Santos Brito, K., & Adeodato, P. J. L. (2020). Predicting Brazilian and US elections with machine learning and social media data. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Enríquez, J. G., Jiménez-Ramírez, A., Domínguez-Mayo, F. J., & García-García, J. A. (2020). Robotic process automation: a scientific and industrial systematic mapping study. *IEEE Access*, *8*, 39113–39129.

Fernandes, J. M., Debus, M., & Bäck, H. (2021). Unpacking the politics of legislative debates. *European Journal of Political Research*, *60*(4), 1032–1045.

Ha, S., & Jeong, H. (2021). Unraveling hidden interactions in complex systems with deep learning. *Scientific Reports*, *11*(1), 12804.

Hegazy, I. M. (2021). The effect of political neuromarketing 2.0 on election outcomes: The case of Trump's presidential campaign 2016. *Review of Economics and Political Science*, *6*(3), 235–251.

Hendriks, C. M., Ercan, S. A., & Boswell, J. (2020). *Mending democracy: democratic repair in disconnected times*. Oxford University Press, USA.

Howard, P. N. (2020). *Lie machines: How to save democracy from troll armies, deceitful robots, junk news operations, and political operatives*. Yale University Press.

Howard, P. N., & Hussain, M. M. (2013). *Democracy's fourth wave?: digital media and the Arab Spring*. Oxford University Press.

Huang, J., He, D., Obaidat, M. S., Vijayakumar, P., Luo, M., & Choo, K.-K. R. (2021). The application of the blockchain technology in voting systems: A review. *ACM Computing Surveys (CSUR)*, *54*(3), 1–28.

Johnson, A. R. (2022). *Anthropogenic Influences on the Decline, Restoration, and Eco-Evolutionary Dynamics of Lake Superior's Coaster Brook Trout*. Michigan Technological University.

Lavezzolo, S., Ramiro, L., & FERNÁNDEZ-VÁZQUEZ, P. (2022). Technocratic attitudes in COVID-19 times: Change and preference over types of experts. *European Journal of Political Research*, *61*(4), 1123–1142.

Li, L., & Zhang, J. (2021). Research and analysis of an enterprise E-commerce marketing system under the big data environment. *Journal of Organizational and End User Computing (JOEUC)*, *33*(6), 1–19.

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, *175*, 121390.

Miller, J. D., Woods, L. T., & Kalmbach, J. (2022). The impact of the Covid-19 pandemic in a polarized political system: Lessons from the 2020 election. *Electoral Studies*, *80*, 102548.

Neshenko, N., Nader, C., Bou-Harb, E., & Furht, B. (2020). A survey of methods supporting cyber situational awareness in the context of smart cities. *Journal of Big Data*, *7*(1), 1–41.

Nie, X., Qin, D., Zhou, X., Duo, H., Hao, Y., Li, B., & Liang, G. (2023). Clustering ensemble in scRNA-seq data analysis: Methods, applications and challenges. *Computers in Biology and Medicine*, 106939.

Rodrigues, A. A., Brito, D. R. B., Kono, I. S., Reis, S. S., & … (2022). Seroprevalence of Neospora caninum and risk factors associated with infection in water buffaloes (Bubalus bubalis) from Maranhão State, Brazil. *…: Regional Studies and …*. https://www.sciencedirect.com/science/article/pii/S2405939021001337

Schuster, S. S. (2020). Does campaign spending affect election outcomes? New evidence from transaction-level disbursement data. *The Journal of Politics*, *82*(4), 1502–1515.

Wang, Y., Sun, Y., Huang, Y., Liu, Z., Gao, S., Zhang, W., Ge, W., & Zhang, W. (2022). Ferv39k: A large-scale multi-scene dataset for facial expression recognition in videos. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20922–20931.

Xavier, D. R., e Silva, E. L., Lara, F. A., e Silva, G. R. R., Oliveira, M. F., Gurgel, H., & Barcellos, C. (2022). Involvement of political and socio-economic factors in the spatial and temporal dynamics of COVID-19 outcomes in Brazil: A population-based study. *The Lancet Regional Health–Americas*, *10*.

Zhukov, D., Khvatova, T., Millar, C., & Andrianova, E. (2022). Beyond big data–new techniques for forecasting elections using stochastic models with self-organisation and memory. *Technological Forecasting and Social Change*, *175*, 121425.