UNDERSTANDING THE LOCAL SOCIOECONOMIC DYNAMICS BASED ON THE POLITICAL COMPASS:
AN INTERDISCIPLINARY APPROACH TO THE CONSTRUCTION OF A VERSATILE COMPUTATIONAL MODEL

Dicle Öztürk[a], Seçil Şen[b], Toykan Doğan[c], Ahmet Yapdı[d]
dicle.ozturk@boun.edu.tr, secil.sen@boun.edu.tr, toykan.dogan@boun.edu.tr, ahmet.yapdi@boun.edu.tr
[a] CmpE, PhD
[b] CmpE, PhD
[c] CogSci, Psy, MA
[d] Soc, BA

All authors are affiliated with the institution of Boğaziçi University

May 2023

ABSTRACT

This research endeavor involves the development of a computational framework to investigate, examine, and subsequently forecast the socioeconomic dynamics of societal subgroups, with a particular focus on their placement within the political compass. The 2018 Turkish presidential election results serve as a localized factor within this study. By integrating diverse datasets, including the geographical information of supermarkets catering to distinct target markets obtained from Google Maps, this study aims to comprehensively discern regressive trends between the socioeconomic characteristics of local voters and their voting inclinations. As a result, a strong correlation has been discovered between upscale supermarkets and the preference for a liberal political alliance in voting behavior.

Keywords: Behavioral economics, regression models, socioeconomic segmentation, voting behavior

Introduction

The societal placement on the political compass can be a convenient and standardized tool for understanding the socio-economic landscape of societies. Elections, through the analysis of voting patterns based on demographics like income, education level, employment status, and occupation, can offer valuable insights into the fundamental values, beliefs, and preferences of citizens. These aspects might substantially be influenced by their socioeconomic backgrounds. By examining these patterns, we can identify societal divisions and disparities in resource access.

Furthermore, election results, particularly localized ballot outcomes, can shed light on voters' socio-political and socio-economic concerns. These results can be interpreted as significant indicators for making informed policy decisions. Analyzing local ballot results can also serve as a valuable analytical tool for understanding larger socioeconomic trends and changes in public opinion, including areas like immigration, healthcare, economics, and trade. Moreover, electoral outcomes can indicate the level of trust and confidence that citizens have in their political positions, which can have significant implications for social cohesion and stability in a broader sense.

Prior Studies

In Gainsborough's (2005) study, the primary distinctions between demographic and socioeconomic factors are elucidated, revealing interconnected patterns with the disparities between urban and suburban regions. These disparities directly influence the contextual profiles in which individuals reside, encompassing various aspects such as income, education, race, and ethnicity. Gainsborough also addresses the noteworthy influence of divergent demographic profiles in suburban and urban areas on residents' perceptions of political ideologies, dynamics, and their consequential outcomes. This influence extends to their voting behavior in presidential elections. Notably, the heterogeneous demographic composition of urban areas, characterized by a wide spectrum of poverty levels, impacts residents' attitudes towards immigration and government expenditure.

Socioeconomic patterns significantly affect political orientations. Individuals' economic circumstances, such as their income level, employment status, and overall economic stability, often influence their political views and voting behavior (Armutcu & Tan, 2021). Those in lower socioeconomic brackets may favor policies and parties that promise social safety nets, job creation, and wealth redistribution. Conversely, those in higher socioeconomic brackets may lean towards parties and policies that promise lower taxes, business-friendly regulations, and economic liberalization. This dynamic is not absolute, as cultural, historical, and personal factors also play a role. However,

the economic situation is a significant determinant of political orientation, as seen in the aforementioned study where an increase in per capita GDP positively affected the vote rate of the ruling party in Turkey.

In terms of socioeconomic patterns affecting political orientations, it's important to note that socioeconomic status often influences an individual's political beliefs and behaviors. Those with higher socioeconomic status may have different political priorities and access to different resources than those with lower socioeconomic status. (Toros, 2014) For example, individuals with higher income levels may be more likely to support policies that favor economic growth and lower taxes, while those with lower income levels may be more likely to support policies that promote social welfare and income redistribution. Additionally, education, a key component of socioeconomic status, can shape political orientations by influencing individuals' awareness of political issues and their capacity for political participation.

In the specific case of Turkey, as once claimed in the article of Toros (2014), political and personal values are becoming increasingly central to voter behavior due to the individualization of political choices and the influence of voters' likes and dislikes. Toros' study also discusses the left-right division in Turkish politics, noting that this division is deeply rooted in the country's political history and is influenced by various factors, including class and constitutional changes. The study further highlights the role of the media in shaping political discourse and influencing voting behavior.

Despite the existence of studies that have identified significant statistical correlations or relationships between socioeconomic dynamics and individuals' political stances, a comprehensive investigation to develop a computational model for analyzing and predicting the relationship between socioeconomic patterns and voting behavior has yet to be undertaken.

Data

In this study, we have utilized a comprehensive dataset comprising the local voting results of Istanbul specifically for the 2018 presidential election in Turkey. The dataset was obtained from the Supreme Election Council (YSK) and provides detailed information on all the ballots cast, as well as the residency streets of the registered voters. By employing this dataset, we aimed to gain insights into the electoral dynamics and patterns within Istanbul during the aforementioned presidential election. The dataset's extensive coverage and its affiliation with the YSK enhance its reliability and accuracy, enabling us to draw robust conclusions and make informed analyses regarding the voting behavior and preferences of Istanbul's electorate during this significant political event.

In conjunction with the primary ballot results data, this study incorporates supplementary geographical data concerning

the presence and locations of various supermarkets, convenience stores, and dollar stores within Istanbul. This additional dataset includes precise information about the exact positions of these establishments, including the streets on which they are situated. The collection of this data was facilitated by the development and implementation of an automated algorithm utilizing the Google API. This algorithm offers a high degree of flexibility, enabling it to efficiently retrieve and process the required information. Moreover, the accuracy of the algorithm ensures that the geographical data obtained is reliable and can be effectively utilized in conjunction with the primary dataset. By integrating these diverse datasets, this research aims to explore potential relationships between the presence of retail establishments and the voting patterns observed in Istanbul during the 2018 presidential election.

Methodology

The utilization of various market types, categorized based on the socioeconomic levels of their target audiences, holds promise as an effective means to assess the geographical dispersion of individuals and their corresponding socioeconomic segments. This approach acknowledges the inherent connection between consumer behavior and socioeconomic factors, indicating that individuals with similar socioeconomic backgrounds often exhibit similar purchasing patterns and preferences.

Scholarly studies, such as Amenta et al. (2010) and Bartels (2016), substantiate the influence of socioeconomic factors on consumer behavior, encompassing choices of products, brand preferences, and shopping habits. Furthermore, investigations exploring the spatial distribution of different market types have revealed that the presence and concentration of specific market segments can act as indicators for the socioeconomic composition of a particular area. By utilizing this approach, researchers can acquire valuable insights into the socioeconomic characteristics and spatial arrangement of individuals, leading to a more comprehensive understanding of the social and economic dynamics within a region. This methodology bridges consumer behavior research and geographical analysis, offering a nuanced perspective on the relationship between individuals, their socioeconomic status, and their spatial distribution.

To investigate the potential correlation between individuals' voting behaviors and their socioeconomic patterns, the research employed a "feature analysis" approach. This analysis aimed to identify the key features that significantly influenced the overall outcome of voting preferences. To achieve this, regression analysis was conducted using the random forest method. The random forest method, as a powerful machine learning technique, offers several advantages for this study. Firstly, it can handle a large number of input variables, allowing for the inclusion of various socioeconomic factors that may impact voting behaviors. Secondly, the random forest method provides a measure of feature importance, which assists in

identifying the specific features that have the most significant impact on voting preferences. This allows us to gain insights into the socioeconomic factors that are particularly influential in shaping political decision-making. Furthermore, the random forest method helps mitigate overfitting issues by introducing randomness in the selection of data and features, leading to more robust and reliable predictions. By employing the random forest method in the feature analysis and regression, this research can uncover the complex relationship between individuals' voting behaviors and their socioeconomic patterns, offering valuable insights into the underlying mechanisms that drive political decision-making processes.

In our previous analysis of the model, it was determined that the Random Forest Regressor demonstrated the most superior performance. To further enhance our comprehension of the factors influencing this performance, we carried out a SHAP analysis. SHAP, an abbreviation for SHapley Additive exPlanations, is an explainable artificial intelligence technique that has garnered considerable popularity. Its effectiveness stems from the Shapley concept, a fundamental principle in cooperative game theory that assigns a value to each participant in a cooperative game based on their incremental contributions. SHAP (Shapley Additive Explanations) serves as a comprehensive framework that leverages Shapley values to explicate the output of any machine learning model. By employing SHAP, we are provided with a means to examine a model by quantifying the contribution made by each feature in predicting the model's output for a specific instance. This analytical approach aids in comprehending the significance and influence of individual features on the model's predictions.

Our objective is to establish a correlation between data from supermarkets and political preferences. To achieve this, we employ a predictive approach to determine the preferred "*ittifak*" (political alliance) of a given region based on supermarket statistics. Our focus lies on two primary alliances, namely the "*millet ittifakı*" and the "c*umhur ittifakı*", which can be classified as a binary outcome. This classification serves as a straightforward means to showcase the efficacy of the random forest model in analyzing political dynamics.

Results

We conducted a pair of experiments to explore specific facets of our research. The initial experiment focused on assessing the impact of the number of supermarkets on the regression performance. Subsequently, our second experiment sought to ascertain the potential fluctuations in the model's performance across diverse neighborhoods.

Experiment 1 - Feature Analysis:

We have three feature sets: 1) Election results for the presidency (CB), 2) Election results for parliament (MV), 3) Number of supermarkets (MRKT); with separated features in each, with their corresponding

values for each street in neighborhoods in Kadıköy, Istanbul.

In this experiment, we run regression on our dataset using the following features:

i) from CB and MRKT set
ii) from only the CB set
iii) from only the MRKT set,
iv) from MV and MRKT set,
v) from only MV set

To predict separately each one of the targets from the following set that includes party names, presidential candidate names, supermarket names, and the number of invalid votes, which we considered to be important factors in making or demonstrating decisions for our specific community:

```
       target_cols = ['itt_mv_CUMHUR
İTTİFAKI', 'itt_mv_MİLLET İTTİFAKI',
                      "pol_mv_HDP",
                      'pol_cb_GEÇERSİZ OY
TOPLAMI',
                      'pol_mv_GEÇERSİZ OY
TOPLAMI',
                      'pol_cb_RECEP TAYYİP
ERDOĞAN', "pol_mv_AK PARTİ",
                      "pol_mv_HÜDA PAR",
                      'pol_cb_MUHARREM
İNCE', "pol_mv_CHP",
                      'mrkt_n_A101',
'mrkt_n_Macro Center', 'mrkt_n_BİM',
                      'mrkt_n_imam hatip
lisesi', 'mrkt_n_ŞOK', 'mrkt_n_migros',
                      'pol_cb_DOĞU
PERİNÇEK', "pol_mv_VATAN PARTİSİ",
                      "pol_mv_İYİ PARTİ",
"pol_cb_MERAL AKŞENER",
                      "pol_mv_SAADET",
"pol_cb_TEMEL KARAMOLLAOĞLU"

                      ]
```

We used a Random Forest regression algorithm with parameters that we determined after a preliminary set of experiments. The results are based on 5-fold cross-validation. The following metrics are used for the aforementioned purpose:

a. ."mean_absolute_error"
b. "mean_squared_error”
c. "median_absolute_error"

These metrics allow an easier interpretation compared to the other regression-related metrics. The negative values are due to the tool we are using, such that it requires applying these metrics negatively in cross-validation. So, the larger the value, the better the performance.

We need to add that the values and the differences between them are quite small. Still, we would like to present our findings in a way that they present some clues regarding the real conditions and are indeed in line with the real extremities of some of the features. The scoring metric we present is a mean absolute error (MAE) (normalized over five folds).

1) Target-wise Analysis:

a) The first five targets predicted with the best performance (more easily predicted)

target	MAE	feature-set
HÜDA PAR	-0.000211625199673209 02	MV
DOĞU PERİNÇEK	-0.000391609388555774 6	MV
VATAN PARTİSİ	-0.000453399907092598	MV
CUMHUR İTTİFAKI	-0.000476155174349498 5	CB
TEMEL KARAMOLLAOĞLU	-0.000534221185283504 3	MV

Table 1: T-W Analysis (a)

As expected, the least easily predictable targets are from the supermarket numbers feature sets.

b) The Effect of the Supermarkets Feature Set on Performance

There is a negligibly small effect of this feature set on learning performance. For almost all targets, using only MV or only CB feature sets gives a (very) slightly better performance. We noticed that MRKT affects targets that have centrist-moderationist inclinations.

2) Location-wise Analysis:

In this part, we partition the data into neighborhoods such that for each neighborhood, we hold out its instances as the test set, and the rest as the train set. We train and test the regression algorithm with this setting (without cross-validation, therefore) and obtain the performance of the algorithm on each neighborhood separately, to see if there is any difference, if there are any fluctuations over these localizations.

The performance metric we present is a mean absolute error (MAE). As in the previous part, the differences in scores are very small but we thought they might be informative.

i) The Neighborhoods with the Best Scores

We present the first five neighborhoods with the target column and MAE score of the corresponding regression model.

neighborhood	target	MAE score
Feneryolu	HÜDA PAR	0.000109091429300 2466
Erenköy	HÜDA PAR	0.000113394638841 40959
Suadiye	HÜDA PAR	0.000150535264308 60683
Suadiye	MİLLET-İTTİFAKI	0.000161018266870 47415
Caferağa	HÜDA PAR	0.000164082371026 15784

Table 2: L-W Analysis (i)

Predicting targets from the MRKT feature set is again remarkably worse performing.

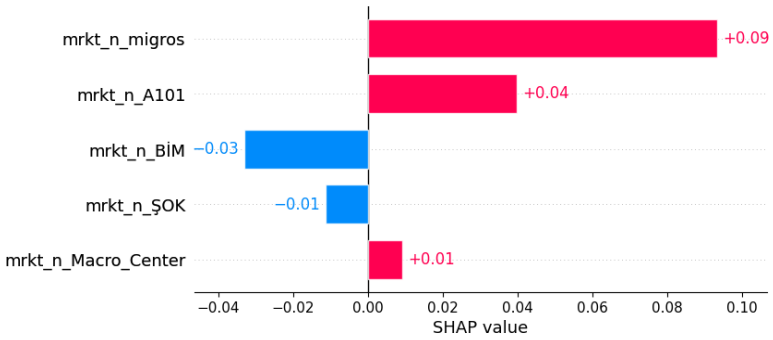
ii) The Most Performant Targets across Neighborhoods

As in the results of the previous experiments, prediction of HÜDA PAR, VATAN PARTİSİ, DOĞU PERİNÇEK and one of the İTTİFAK groups (which is elected by the most attentive) are the targets where the model performs the best.

Experiment 2 - SHAP Analysis

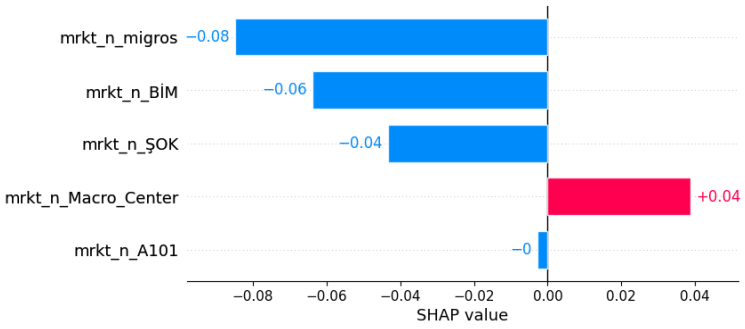
The initial results of our study reveal a relatively high performance in this classification task, as evidenced by an F1 score of 0.89 on the test dataset. To gain deeper insights into the model, we conducted an analysis utilizing SHAP (Shapley Additive Explanations), which enabled us to determine the importance of each indicator corresponding to the supermarket statistics. These weights provide valuable information regarding the significance of different indicators in predicting the political preferences of a region.

Graph 1 provides a visual representation of the impact of various indicators on the preference for "cumhur ittifakı" over "millet ittifakı". It demonstrates that selections favoring "cumhur ittifakı" are positively influenced by the numbers associated with Migros, A101, and Macro Center. On the other hand, they are negatively influenced by the numbers associated with Bim and Şok. This suggests that as the presence and prominence of Bim or Şok increase in a particular region, there is a higher likelihood that the region will not choose "cumhur ittifakı" and instead opt for "millet ittifakı", and vice versa.



Graph 1: SHAP computation. on cumhur over millet

Graph 2 illustrates the impact of various indicators in the selection of "millet ittifakı" (national alliance) over "cumhur ittifakı" (people's alliance). The preferences for "millet ittifakı" are positively influenced by the numerical values associated with Macro Center alone, while being negatively influenced by the numerical values associated with Bim, Şok, Migros, and A101. This suggests that when the presence or size of Bim or Migros increases in a specific region, the likelihood of that region choosing "millet ittifakı" decreases, leading them to opt for "cumhur ittifakı" instead, and vice versa.

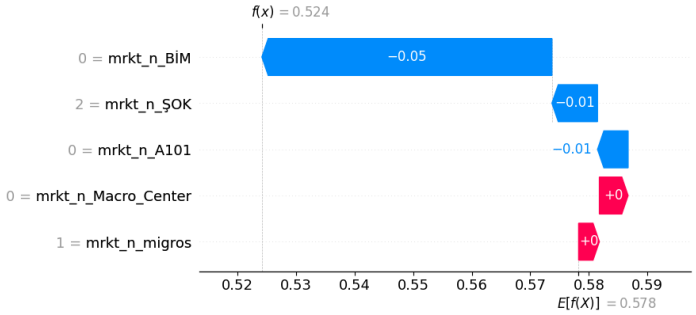


Graph 2: SHAP computation on millet over cumhur

Based on the provided findings, it can be inferred that Macrocenter, a supermarket catering to the higher socioeconomic stratum, is a significant indicator for the millet ittifakı (national alliance). Conversely, Migros and A101 at the local level demonstrate statistically significant associations with voting preferences towards cumhur ittifakı (people's alliance). Nevertheless, our current dataset has limited breadth, as it is confined to the geographical boundaries of Istanbul province. Consequently, the excessive concentration of Migros and A101 stores in Istanbul may impact the significance of our regression model.

Therefore, it is strongly recommended that future research expands the model to larger datasets encompassing diverse provinces and cities, where the distribution of the aforementioned supermarkets is not uniformly dispersed. The influence of Macrocenter, however, reinforces our assertion regarding the socioeconomic influences on voting preferences.

The depicted Graph 3 provides an overview of how indicators affect the final prediction when considering the sizes of supermarkets, with the "millet ittifakı" alliance as the ground truth. The model generates a scalar value of 0.524, representing the predicted probability of belonging to the "millet ittifakı" alliance. Notably, the size of BİM exhibits the most significant negative influence on the model's prediction. While the other indicators also contribute to the prediction, their impact is not as prominent as that of BİM size. Additionally, it is important to note that E[f(x)] refers to the expectation within this particular context.



Graph 3: Final computation of indicators

Discussion

This study was meticulously constructed using data about elections at the district and neighborhood levels within the Kadıköy district. As demonstrated in the 2023 Turkish General Elections, the necessity for parties to secure data flow at polling stations has become increasingly crucial. To enhance the accuracy of nationwide election outcome predictions, it is suggested that emphasis be placed on conducting analogous studies across all regions within Turkey. Future investigations could extrapolate from this

study, utilizing data obtained from the Supreme Election Council (YSK), the Ministry, and political factions.

This investigation endeavored to decipher the correlation between market dynamics and electoral preferences. It is paramount to meticulously scrutinize neighborhoods and residential zones from diverse perspectives. To formulate precise conclusions germane to the study area, it is beneficial to calculate the influence wielded by all regional public and private institutions on electoral outcomes. Furthermore, the income bracket of the district's residents served as a substantial indicator in forecasting electoral outcomes. The architectural and urban structure of the neighborhoods could be augmented by sociological analyses to refine these predictions. For instance, street graffiti may be indicative of ghettoization within the region, and the inclusion of such minutiae in the computations could yield more precise results. The study's efficacy could be enhanced through the incorporation of municipal data. A broader range of detailed data from municipalities, such as parks, parking facilities, art centers, and libraries could be harnessed for these enhancements. Concurrently, images of the streets could be processed through a visual processing model to determine the region's developmental index, further contributing to a more nuanced understanding of the electoral landscape.

While economic income levels within a region can sometimes provide valuable insight for predicting elections, they can also prove misleading. Economic crises do not always impact general elections in a manner proportional to the severity of the crisis. Parameters beyond economic ones such as religious beliefs, racial and age demographics, the social identities of communities, the period activities of parties in the region, the local socio-political structures, and the influence of media, all play a significant role in the forecasting of election outcomes.

The integration of these additional factors into the predictive model would be beneficial. The religious, racial, and age composition of a district, the social identities of its communities, the activities and initiatives of political parties during different periods, the underlying socio-political structures, and the influence of media – all these aspects constitute influential factors. Their inclusion in predictive models can enrich our understanding of electoral behavior and enhance the accuracy of election outcome forecasts. As such, predictive models would be well served by taking a multi-faceted approach that accounts for these diverse influences.

References and Other Annotations

Amenta, E., Caren, N., Chiarello, E., & Su, Y. (2010). The political consequences of social movements. Annual Review of Sociology, 36, 287-307.

Bartels, L. M. (2016). Unequal democracy. In Unequal Democracy. Princeton University Press.

Carmines, E. G., & Stimson, J. A. (1980). The two faces of issue voting. American Political Science Review, 74(1), 78-91.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1980). The American voter. University of Chicago Press.

Gainsborough, J. F. (2005). Voters in context: Cities, suburbs, and presidential vote. American Politics Research, 33(3), 435-461.

Google. (n.d.). Google Maps. Retrieved April 20, 2023, from https://maps.google.com/

Karnig, A. K. (1976). Black representation on city councils: The impact of district elections and socioeconomic factors. Urban Affairs Quarterly, 12(2), 223-242.

Manza, J., & Brooks, C. (1999). Social cleavages and political change: Voter alignments and US party coalitions. OUP Oxford.

Newman, B. I., & Sheth, J. N. (1985). A model of primary voter behavior. Journal of Consumer Research, 12(2), 178-187.

Open Data - Istanbul Metropolitan Municipality. (2023, April). Istanbul. Retrieved April 20, 2023, from https://data.ibb.gov.tr/en/.

Peterson, M., & Epler, R. (2020). Sustainability developments in cities of the world. Continuing to Broaden the Marketing Concept, 17, 243-260.

Sartori, G. (1969). From the sociology of politics to political sociology. Government and Opposition, 4(2), 195-214.

Stellar, J. E., Manzo, V. M., Kraus, M. W., & Keltner, D. (2012). Class and compassion: socioeconomic factors predict responses to suffering. Emotion, 12(3), 449

Toros, E. (2014). Social indicators and voting: The Turkish case. Social indicators research, 115, 1011-1029.

Van Prooijen, J. W., Krouwel, A. P., Boiten, M., & Eendebak, L. (2015). Fear among the extremes: How political ideology predicts negative emotions and outgroup derogation. Personality and social psychology bulletin, 41(4), 485-497.