ПОЛИТОЛОГИ XXIV
Эрдэм шинжилгээний бичиг №24: 167-182 дох тал

PREDICTING VOTER CHOICES: THE CASE OF MONGOLIA

Zolzaya Gungaa
Department of Statistics and Econometrics,
School of Economics and Business, MULS, Mongolia

Usukhbayar Bayarsaikhan
Department of Computer Science,
School of Applied Sciences, MULS, Mongolia

Abstract: In democratic societies, the electoral process constitutes the primary mechanism for the legitimate formation of political authority, with voters' political choices directly influencing socio-economic development and policy outcomes. Research on voter behavior in Mongolia remains limited, with most studies relying on descriptive statistical summaries. This study employs a Multinomial Logistic Regression (MLR) model to analyze the combined effects of demographic characteristics, socio-economic indicators, and party evaluations among voters from two electoral districts. The results indicate that age, education, income, sources of political information, political activism, previous voting behavior, and party evaluations are statistically significant predictors of voter choice ($p < 0.05$). The model accurately predicts support for the Mongolian People's Party (MPP) and Democratic Party (DP) with 84–98% precision, whereas predictions for the 'Other' category are weaker, likely due to heterogeneity and a small number of observations. Model fit indices (Cox & Snell $R^2 = 0.745–0.791$; Nagelkerke $R^2 = 0.803–0.852$) demonstrate a high explanatory capacity. The findings underscore the theoretical and methodological relevance of data-driven electoral forecasting in the Mongolian context and highlight the need for further research into small-party heterogeneity and latent voter structures.

Keywords: Multinomial logistic regression for qualitative data, Electoral forecasting

Introduction

In democratic societies, the electoral process serves as the fundamental mechanism for the legitimate formation of political authority, with voters' political choices directly influencing national development, policymaking, and socio-economic trends [1], [2]. Consequently, predicting voter behavior has emerged as a critical research area, not only in political science but also in the fields of information technology, statistics, and data analytics[3].

Since Mongolia's transition to a democratic system, multiple parliamentary, presidential, and local elections have been conducted. The outcomes of these elections have varied according to numerous factors, including societal sentiment, economic conditions, political party reputations, mass media, and social networks [4], [5]. However, scientific studies that systematically analyze and predict actual voter behavior remain relatively scarce in the Mongolian context [6].

Internationally, studies examining determinants of voter behavior frequently consider variables such as age, gender, education, income, geographic location, and political ideology[7], [8]. For instance, research conducted in the United States and Europe has demonstrated that machine learning algorithms—including logistic regression, decision trees, neural networks, and support vector machines (SVMs)—can reliably forecast electoral outcomes[9], [10]. Furthermore, sentiment analysis of social media data (e.g., Twitter, Facebook) has enabled real-time detection of shifts in voter preferences[11], [12].

---

[1] A.Blais "To vote or not to vote: The merits and limits of rational choice theory", 2000

[2] P.Norris "Why electoral integrity matters", Cambridge University Press, 2014

[3] S.Gueuwou, R.-M.Mensah "Stance prediction and analysis of Twitter data: A case study of Ghana 2020 presidential elections", arXiv. Retrieved from https://arxiv.org/abs/2306.14203, 2023

[4] T.Batbold, D.Tumenjargal "Factors influencing voter participation in Mongolia", Politology, 18(2), 45-58, 2018

[5] N.Myagmartsooj "Factors influencing voter turnout in Mongolia (Youth case study)", Politology, 22(1), 111–118, 2024

[6] G.Urandelger, S.Burmaa, M.Altan-Od "Factors influencing youth political choice in Mongolia", Journal of Business and Innovation, 9(2), 55–70, 2023

[7] P.Norris "Why electoral integrity matters", Cambridge University Press, 2014

[8] G.Shahi, A.Basyurt, S.Stieglitz "Agenda formation and prediction of voting tendencies for European Parliament election using textual, social and network features", Information Systems Frontiers. doi:10.1007/s10796-024-10568-w, 2025

[9] K.Brito, P.Adeodato "Machine learning for predicting elections in Latin America based on social media engagement and polls", Government Information Quarterly, 40(1), 101782. doi:10.1016/j.giq.2022.101782, 2023

[10] S.Gueuwou, R.-M.Mensah "Stance prediction and analysis of Twitter data: A case study of Ghana 2020 presidential elections", arXiv. Retrieved from https://arxiv.org/abs/2306.14203, 2023

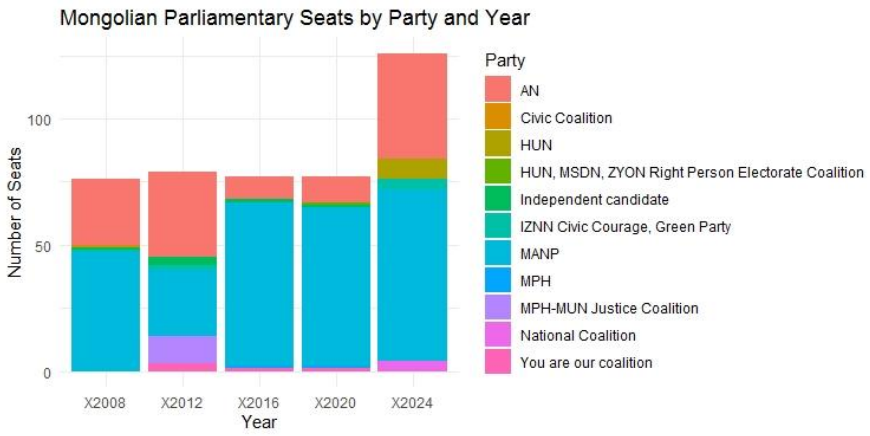
[11] G.Shahi, A.Basyurt, S.Stieglitz "Agenda formation and prediction of voting tendencies for European Parliament election using textual, social and network features", Information Systems Frontiers. doi:10.1007/s10796-024-10568-w, 2025

[12] S.Gueuwou, R.-M.Mensah "Stance prediction and analysis of Twitter data: A case study of Ghana 2020 presidential elections", arXiv. Retrieved from https://arxiv.org/abs/2306.14203, 2023

Global organizations, including the World Bank and the International Foundation for Electoral Systems (IFES), regularly conduct statistical analyses of electoral processes, voter participation, and political behavior, highlighting the importance of data-driven insights [13], [14]. Such studies consistently identify macro-level factors—economic conditions, trust in government, corruption levels, and media influence—as significant determinants of voter choice [15].

Between 2008 and 2024, the number of seats held by political parties fluctuated steadily, with abrupt changes occurring at specific intervals, indicating that the institutional consolidation of parliamentary representation has not yet been fully established. To provide a comprehensive depiction of the evolving dynamics within the party competition landscape, the following graph has been presented.

Figure 1. Seats won by parties, by year



Source: NSO

This graph illustrates the dynamic changes in parliamentary representation of Mongolian political parties between 2008 and 2024. While the Mongolian People's Party (MPP) maintained a stable dominant position, the abrupt fluctuations in the seats of the Democratic Party (DP) reflect institutional instability. During the period under study, newly emerging coalitions and smaller parties began to secure representation to some extent, indicating an increasing level of fragmentation within the party system. The results of 2024 show a sharp departure from previous cycles, revealing new organizational arrangements in political competition and shifts in voter behavior.

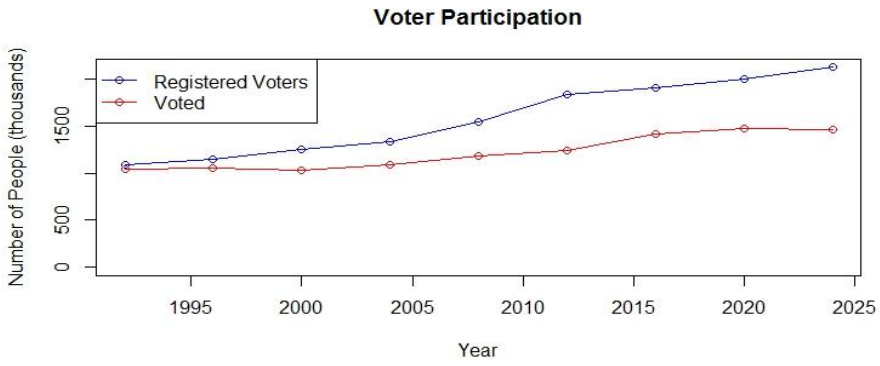
[13] W.Bank "Worldwide governance indicators and voter behavior analysis", World Bank Publications, 2022

[14] International Foundation for Electoral Systems "Voter participation and public trust in elections", IFES, 2021

[15] W.Bank "Worldwide governance indicators and voter behavior analysis", World Bank Publications, 2022

Using voter indicators from the parliamentary elections, the graph also depicts trends in voter preferences. It clearly shows the tendency of the electorate to direct support toward either the ruling party or the opposition.

Figure 2. Number of voters, number of citizens who voted, by year
Source: NSO



Over the period under study, the number of citizens registered on the voter lists steadily increased, whereas the growth in actual voters was comparatively slower, resulting in a persistent gap in electoral participation. Accordingly, this graph was developed and utilized to empirically illustrate the structure of voter turnout and its long-term trends.

In Mongolia, sociological studies have addressed aspects of voter behavior and electoral participation. The National Statistics Office regularly publishes reports correlating election-age citizens' participation with socio-economic indicators, while surveys by the Sant Maral Foundation provide insights into voters' political attitudes [16], [17]. Nevertheless, most of these analyses are descriptive and rely on statistical summaries, with relatively few studies employing predictive models based on machine learning or data mining techniques[18].

With recent advancements in Mongolia's information technology sector, attempts to process electoral data algorithmically and generate predictive models have emerged. Yet these efforts remain largely experimental, constrained by limited access to high-quality datasets. Moreover, studies leveraging social media data

[16] National Statistics Office of Mongolia "Population and election participation statistics", Ulaanbaatar, 2023

[17] Sant Maral Foundation "Public opinion survey on political participation in Mongolia", Ulaanbaatar, 2022

[18] N.Myagmartsooj "Factors influencing voter turnout in Mongolia (Youth case study)", Politology, 22(1), 111–118, 2024

are still in developmental stages, with ethical and privacy considerations unresolved[19].

Overall, while international methodologies for accurately predicting voter choices are relatively mature, Mongolia lacks sufficiently comprehensive official statistical surveys and sample-based studies, highlighting a clear need to advance both theoretical and applied research in this domain[20], [21].

Materials and method

Numerous theories and models have been developed to understand the factors influencing voter choice, which are crucial for examining the interplay between voter behavior and socio-economic conditions[22]. According to Rational Choice Theory, voters determine which party or candidate to support based on the maximization of personal utility[23]. This theory conceptualizes voters as deliberate and calculative actors, making it widely applicable in statistical models designed to predict electoral outcomes [24]. Sociological models emphasize that social factors such as age, gender, education, income, occupation, and geographic location significantly shape voting behavior [25]. These models allow for a nuanced understanding of voters that integrates both social context and cultural environment alongside rational calculation.

Psychological models highlight the influence of attitudinal factors—including party loyalty, political activism, and information-seeking habits—on voter decision-making [26]. Economic-choice theories posit that macroeconomic conditions at the time of elections, such as unemployment, inflation, and living standards, can directly affect voter decisions. In the Mongolian context, these theoretical frameworks collectively indicate that predictive analyses of voter behavior must account simultaneously for social, economic, and informational determinants.

Previous studies in Mongolia have consistently shown that voters' education level, age group, prior voting behavior, and information sources exert notable

[19] S.Gueuwou, R.-M.Mensah "Stance prediction and analysis of Twitter data: A case study of Ghana 2020 presidential elections", arXiv. Retrieved from https://arxiv.org/abs/2306.14203, 2023

[20] P.Norris "Why electoral integrity matters", Cambridge University Press, 2014

[21] G.Urandelger, S.Burmaa, M.Altan-Od "Factors influencing youth political choice in Mongolia", Journal of Business and Innovation, 9(2), 55–70, 2023

[22] A.Blais "To vote or not to vote: The merits and limits of rational choice theory", 2000

[23] A.Downs "An economic theory of democracy", 1957

[24] B.Hair, A.Babin "Multivariate data analysis", Cengage, 2019

[25] N.Myagmartsooj "Factors influencing voter turnout in Mongolia (Youth case study)", Politology, 22(1), 111–118, 2024

[26] G.Urandelger, S.Burmaa, M.Altan-Od "Factors influencing youth political choice in Mongolia", Journal of Business and Innovation, 9(2), 55–70, 2023

influence on electoral choices. Building on these insights, the present study integrates the aforementioned theoretical perspectives to develop a model capable of predicting voter choices in Mongolia. The study employs a Multinomial Logistic Regression (MLR) framework, which accommodates dependent variables with multiple categories [27]. Here, the dependent variable is the political party chosen by the voter, while independent variables include age, gender, education, income, information sources, political activism, and prior voting behavior. Regional differences and other potential control variables are also incorporated into the model.

Data collection was conducted through a structured survey designed to capture voters' socio-economic characteristics, political engagement, and information-seeking patterns [28]. Survey items were measured using nominal and ordinal scales, with Likert scales employed where appropriate to quantify attitudinal and behavioral indicators [29].

For the Multinomial Logistic Regression (MLR) model, the following formulation is employed:

$$Logit\left(\frac{P(Y=i)}{P(Y=base)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n \qquad (1)$$

In this context, YYY represents the voter's choice, while $X1, \ldots, Xn$ denote the independent variables, and β\betaβ represents the model coefficients. To assess model adequacy, indicators such as the Likelihood Ratio Test, Pseudo R^2, and Classification Accuracy are employed to evaluate the predictive validity of the model. Additionally, checks for multicollinearity and other assumptions necessary for reliable estimation were conducted to ensure the robustness of the results [30].

The study sample comprises 138 voters from two electoral districts (Gobi-Altai, Zavkhan, Uvs, and Khovd) and 124 voters from four districts (Bulgan, Orkhon, and Khuvsgul), ensuring representation of major voter populations in Mongolia. The sample size was determined to be sufficient to provide stable and reliable estimates for the Multinomial Logistic Regression (MLR) model.

The survey encompassed 138 voters from the 2nd electoral district (Gobi-Altai, Zavkhan, Uvs, Khovd) and 124 voters from the 4th electoral district (Bulgan, Orkhon, Khuvsgul), yielding a total sample of 262 respondents. Following established guidelines for multinomial logistic regression [31], the required minimum sample size to achieve stable parameter estimates is computed as $N \geq$

[27] B.Hair, A.Babin "Multivariate data analysis", Cengage, 2019

[28] National Statistics Office of Mongolia "Population and election participation statistics", Ulaanbaatar, 2023

[29] B.Hair, A.Babin "Multivariate data analysis", Cengage, 2019

[30] L.Hosmer "Applied logistic regression", Wiley, 2013

[31] L.Hosmer "Applied logistic regression", Wiley, 2013

$10 \times (k-1) \times p = 330$, where k is the number of outcome categories and p the number of predictors. Although the total sample slightly undercuts this threshold, the distribution of respondents across major political parties and predictors ensures sufficient data to generate reliable and interpretable MLR estimates.

The primary data for this study were collected to capture voters' socio-economic characteristics, political attitudes, and behavioral patterns. Data collection employed a structured survey designed to be representative of Mongolia's adult voting population. The survey comprised the following sections:

1. **Demographic Information:** Socio-economic variables measured included age, gender, education level, and average monthly household income.

- *Age* was treated as a continuous variable but could also be categorized into age groups for analysis.
- *Gender* was recorded as a binary categorical variable (male, female).
- *Education level* was an ordinal variable with the following categories: Primary, Secondary, Upper Secondary, Diploma, Bachelor, Master, and Doctorate.
- *Average monthly household income* was also treated as an ordinal variable, categorized as: below 1 million MNT, 1–2 million, 2–3 million, 3–5 million, and above 5 million MNT.

2. **Political Participation and Attitudes:** To predict electoral behavior and party choice, the following indicators were employed:

- *Information sources:* Voters were asked where they primarily obtain political information. Categories included television, Facebook, TikTok, news websites, acquaintances/friends, and others. This was treated as a nominal variable.
- *Political activism:* Voters self-assessed their level of political engagement on a 5-point Likert scale, ranging from "Very inactive" to "Very active," treated as an ordinal variable.
- *Previous electoral participation:* Respondents indicated whether they voted in prior parliamentary elections, with response options: Yes, no, or don't remember. This was treated as a nominal variable.
- *Party evaluation:* Voters rated the policies and performance of political parties on a scale from "Very poor" to "Very good," treated as an ordinal variable. Party evaluation was considered a key predictor influencing trust in, and likelihood of supporting, a given party.
- *Current choice:* Respondents indicated which party they would vote for if an election were held today. This categorical (nominal) variable served as

the dependent variable in the study and was used in the Multinomial Logistic Regression (MLR) model.

RESULT

The reliability and validity of the survey instruments administered to participants were thoroughly assessed to ensure the accuracy and credibility of the collected data.

Table 1. Results of reliability and plausibility analysis

Test	2nd circle	4th circle
Cronbach's alpha	0.417	0.441
KMO	0.839	0.658

The results indicate that Cronbach's alpha values were 0.417 for the 2nd district and 0.441 for the 4th district, suggesting relatively low internal consistency within these item sets. This implies that some items may not perfectly correlate with one another or fully represent the intended construct, highlighting the need to further assess the extent to which the included items capture coherent concepts and underlying characteristics.

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.839 for the 2nd district and 0.658 for the 4th district, indicating that the sample sizes were sufficient to perform factor analysis. Specifically, a KMO value around 0.8 is generally considered "meritorious," while a value around 0.6 is regarded as "mediocre but adequate."

To evaluate the internal consistency of the measurement scales, item-level analyses based on Corrected Item–Total Correlation (CITC) were conducted. This approach assesses the extent to which each item reflects the overall scale. Items with CITC ≥ 0.30 are deemed consistent with the general structure of the scale, whereas items with CITC < 0.30 indicate weak structural alignment and may require removal, re-specification, or reassignment to a different factor.

Table 2. Item-Total correlation test results

Variables	2nd circle	4th circle
Age	0.587	0.678
Gender	0.498	0.433
Education	0.853	0.917
Income	0.792	0.798

Source of political information	**-0.248**	-0.315
Political activity	0.765	0.796
Voted in previous elections	**-0.178**	**-0.214**
MPP rating	0.718	0.673
DP rating	0.506	0.529
HUN rating	-0.588	-0.571
Other party ratings	0.346	0.409
Party to vote for today	-0.477	-0.384

The analysis indicates that the internal consistency of the items included in the survey varies across districts. Variables such as Age, Gender, Education (Edu), Income, Political Activism (Activism), and AN_rate exhibited Corrected Item–Total Correlations (CITC) above 0.30 in both districts, suggesting that these items consistently and reliably contribute to the overall measurement structure. Notably, Education and Income demonstrated particularly high CITC values (0.79–0.92), indicating strong alignment with the latent constructs measured in the survey.
Conversely, items such as Information Source (Infosource), Previous Vote (PrevVote), and HUN_rate in the 2nd district, as well as Other_rate and VoteToday in the 2nd district, exhibited CITC values below 0.30, with some instances showing negative correlations. These findings imply that these items may not adequately reflect the intended constructs and may require further examination, potential revision, or reassignment to improve the coherence of the measurement scales.

Table 3. Demographic Characteristics of the Participants

Indicator	**2nd circle**		**4th circle**	
	frequency	Percentage	frequency	Percentage
Age				
Up to 30	24	17.4	19	15.3
31-40	54	39.1	49	39.5
41-50	50	36.3	46	37.1
51-60	9	6.5	10	8,1
61 and up	1	0.7	0	0

Total	138	100	124	100
Gender				
Male	69	50	63	50.8
Female	69	50	61	49.2
Total	138	100	124	100
Education				
Primary	15	10.9	12	9.7
Secondary	40	29.0	37	29.8
Diploma	23	16.7	21	16.9
Bachelor	33	23.9	29	23.4
Master	27	19.6	25	20.2
Total	138	100	124	100
Average household income				
Up to 1 million tugrug	6	4.3	4	3.2
1-2 million tugrug	35	25.4	33	26.6
2-3 million tugrug	38	27.5	33	26.6
3-5 million tugrug	40	29.0	38	30.6
More than 5 million tugrug	19	13.8	16	12.9
Total	138	100	124	100

Regarding age, the 31–40 and 41–50 age groups accounted for the largest proportions in both districts, approximately 39% and 36–37%, respectively, while individuals aged 61 and above represented a very small fraction of respondents.
In terms of gender, the distribution of male and female participants was roughly balanced in both districts, each comprising approximately 50% of the sample.
Comparing education levels, respondents with upper secondary and bachelor's degrees constituted the majority, while those holding a master's degree represented about 19–20% of the total participants.
With respect to average household income, households earning 2–5 million MNT per month accounted for the largest share, approximately 55%, whereas households earning below 1 million MNT comprised the smallest proportion.
Overall, these results indicate that the demographic structures of the two districts are largely comparable, demonstrating a balanced composition of the survey's baseline characteristics.

Let Y denote the voter's choice, with "Other" as the reference category. The multinomial logit model is specified as:

$$log\frac{P(Y=j)}{P(Y=other)} = \beta_{0j} + \sum_{i=1}^{11}\beta_{ij}X_i, \quad j = \{MAN, AN, HUN\} \tag{2}$$

where X_i represents the individual-level predictors including age, gender, education, income, political information source, activism, previous vote, and party evaluations. The estimated coefficients (β_{ij}) and their significance levels are reported in Table 3.

Table 4. Results of Multinomial Logit Model Estimation

Variables	2nd circle		4th circle	
	Choose MPP	Choose DP	Choose MPP	Choose DP
Age	-0.151* (0.124) [0.674^{L}-1.097^{U}]	-0.396* (0.165) [0.487^{L}-0.930^{U}]	-0.178* (0.187) [0.580^{L}-1.207^{U}]	-0.547* (0.259) [0.348^{L}-0.961^{U}]
Gender	0.040* (1.217) [0.096^{L}-11.304^{U}]	-0.993* (1.066) [0.046^{L}-2.995^{U}]	-1.800* (1.422) [0.010^{L}-2.682^{U}]	-1.348* (1.759) [0.008^{L}-8.165^{U}]
Education	0.412* (1.062) [0.188^{L}-12.120^{U}]	1.044* (0.973) [0.422^{L}-19.122^{U}]	-1.271* (1.077) [0.034^{L}-2.319^{U}]	-0.640* (1.121) [0.059^{L}-4.742^{U}]
Income	0.820* (1.248) [0.197^{L}-26.225^{U}]	-2.444* (1.391) [0.006^{L}-1.326^{U}]	2.263* (1.366) [0.660^{L}-139.826^{U}]	1.117* (1.810) [0.088^{L}-106.080]
Source of political information	-0.275* (0.320) [0.406^{L}-1.421^{U}]	1.162* (0.427) [1.385^{L}-7.382^{U}]	-0.728* (0.310) [0.268^{L}-0.872^{U}]	2.097* (0.885) [1.438^{L}-46.124^{U}]
Political activity	-0.636* (1.324) [0.040^{L}-7.098^{U}]	3.889* (1.739) [1.616^{L}-1478.0^{U}]	-0.170* (1.672) [0.032^{L}-22.341^{U}]	3.900* (2.431) [0.421^{L}-5796.3^{U}]
Voted in previous elections	-1.766* (1.287) [0.014^{L}-2.130^{U}]	1.557* (0.769) [1.051^{L}-21.430^{U}]	-0.857* (0.805) [0.088^{L}-2.057^{U}]	1.070* (0.943) [0.460^{L}-18.501^{U}]
MPP rating	1.092* (0.893) [0.518^{L}-17.171^{U}]	-1.545* (0.915) [0.035^{L}-1.281^{U}]	0.968* (0.695) [0.674^{L}-10.283^{U}]	-0.645* (1.195) [0.050^{L}-5.453^{U}]
DP rating	0.596* (0.547) [0.621^{L}-5.302^{U}]	1.720* (0.721) [1.360^{L}-22.922^{U}]	-0.143* (0.669) [0.234^{L}-3.212^{U}]	0.953* (1.285) [0.209^{L}-32.205^{U}]
HUN rating	-1.264* (0.782) [0.406^{L}-1.309^{U}]	-3.635* (1.022) [0.004^{L}-0.196^{U}]	0.023* (0.657) [0.283^{L}-3.707^{U}]	-2.433* (1.182) [0.009^{L}-0.889^{U}]

Other party ratings	-0.519* (0.725) [0.061^{L}-6.958^{U}]	0.240* (1.035) [0.167^{L}-9.658^{U}]	0.708* (0.925) [0.331^{L}-12.427^{U}]	-2.568* (2.176) [0.001^{L}-5.455^{U}]
Constant	3.352	12.314	5.677	11.024
Note	Log-likelihood = 122.28, p = .000; logistic regression (LR) $\chi 2$ =572.271; Cox and Snell R2 = .791, Nagelkerke R2 = .852, McFadden R2 = .593; *Indicates that it is significant at 5% significance level.		Log-likelihood = 101.34, p = .000; logistic regression (LR) $\chi 2$ = 79.873; Cox and Snell R2 = .745, Nagelkerke R2 = .803, McFadden R2 = .519; *Indicates that it is significant at 5% significance level	

In the SPSS implementation of the multinomial logistic regression model, the "Other" category was automatically designated as the reference category. Consequently, the estimated coefficients for the remaining parties are calculated relative to this reference category. While "Other" itself does not appear directly as a coefficient in the model, it serves as the baseline for interpreting all parameter estimates, which is standard statistical practice.

The relatively high standard error observed for the HUN party coefficient indicates low confidence in the true value of this parameter and reflects model instability. This may result from a small sample size, high correlations among covariates, sparse data, outliers, or potential model misspecification, which is why the coefficient was not included in the table.

The multinomial logistic regression results show that age, gender, education, income, political information sources, political activism, prior voting behavior, and party evaluations have statistically significant effects on support for the MPP (Mongolian People's Party) and DP (Democratic Party) ($p < 0.05$). Coefficients from the 2nd and 4th districts indicate that the effect of each predictor may be positive or negative depending on the voter category and party choice. For example, age exhibited a negative association with MPP and DP support, whereas political activism had a positive effect on DP support, reflecting differences in voter behavior.

Model fit indicators (Cox and Snell $R^2 = 0.745–0.791$, Nagelkerke $R^2 = 0.803–0.852$, McFadden $R^2 = 0.519–0.593$) demonstrate a high explanatory power, and the Likelihood Ratio χ^2 test confirms the model's statistical significance ($p < 0.001$). These findings systematically illustrate how individual characteristics, prior voting behavior, and party evaluations influence support for the MPP and DP.

In the multinomial logit framework, with k outcome categories (in this case, VoteToday = MPP, DP, HUN, Other), one category is selected as the reference. Here, "Other" serves as the reference category (Y = Other).

$$log\frac{P(Y=j)}{P(Y=other)} = \beta_{0j} + \beta_{1j}Age + \beta_{2j}gender + \beta_{3j}edu + \cdots + \beta_{11j}other_rate \quad (3)$$

$P(Y=j)$ - Probability of voting for category j

β_{0j} - intercept (base coefficient)

β_{ij} - the effect of variable i on category j

For example, the equation for calculating the probability of choosing an MPP in the second circle is written as follows:

$$log\frac{P(Y=MPP)}{P(Y=other)} = 3.352 - 0.151Age + 0.040gender + 0.412edu + \cdots + -0.519other_rate$$

Using a multinomial logit model equation, the probability of a party being elected based on the electoral quality data was calculated for each circle.

Table 5. Elected party prediction results

Observed	2nd circle Predicted					4th circle Predicted				
	MPP	DP	HUNP	Other P	Percent Correct	MPP	DP	HUNP	Other P	Percent Correct
MPP	47	2	0	1	94.0%	46	1	0	0	97.9%
DP	2	36	1	4	83.7%	2	38	0	2	90.5%
HUNP	0	0	14	6	70.0%	0	0	20	1	95.2%
Other P	8	3	3	11	44.0%	1	3	7	3	21.4%
Overall Percentage	41.3%	29.7%	13.0%	15.9%	78.3%	39.5%	33.9%	21.8%	4.8%	86.3%

The classification tables from the multinomial logistic regression model indicate that the model predicts voter choices for the MPP and DP with relatively high accuracy (MPP: 94.0%, DP: 83.7%). Prediction accuracy for the HUN party is

70.0%, while the “Other” category achieves only 44.0%. The lower accuracy for smaller or heterogeneous categories reflects the limited number of observations within these groups, greater heterogeneity, and the inherent difficulty in predicting support for minor parties. Overall, the model achieves a total prediction accuracy of 78.3%, demonstrating that it captures the voting behavior for the major parties reasonably well, though predictions for smaller and more heterogeneous categories remain relatively weak.

In another instance, the classification results show even higher prediction accuracy for the major parties (MPP: 97.9%, DP: 90.5%, HUN: 95.2%), indicating that the structure of voter preferences and support for these parties is well represented in the model. In contrast, the “Other” category shows a substantially lower accuracy (21.4%), which can be attributed to the small number of observations, high heterogeneity, and the reference-category-based structure of the model. This suggests the model faces challenges in predicting individual voter behavior for minor parties. Overall, the total classification accuracy reaches 86.3%, confirming that the model can reliably predict voter choices for the main parties with high precision.

CONCLUSION

Empirical studies on data-driven electoral forecasting in Mongolia remain relatively scarce, and most existing analyses are of an exploratory nature. The Multinomial Logistic Regression (MLR) model employed in this study integrates voter behavior, socio-economic indicators, and party evaluations, demonstrating the theoretical and methodological relevance of developing data-driven pre-election predictions in the Mongolian context.

The results indicate that voter choice is directly influenced by age, education, household income, sources of political information, political activism, prior voting behavior, and party evaluations. Demographic analysis shows that the sample composition across the two districts was relatively stable, with the majority aged 31–50 and a balanced gender distribution. In terms of education, respondents with complete secondary, bachelor’s, or master’s degrees predominated, while households earning 2–5 million MNT per month were most represented, indicating a balanced baseline demographic profile.

Internal consistency measures, including Cronbach’s alpha and Corrected Item–Total Correlation (CITC), suggest that while some variables (e.g., Infosource, PrevVote, HUN_rate, Other_rate, VoteToday) may not fully capture the latent construct, Age, Gender, Education, Income, Activism, and party evaluation align

well with the scale. The Kaiser-Meyer-Olkin (KMO) coefficient confirmed adequate sampling for factor analysis.

MLR results demonstrate that age, gender, education, income, political information sources, political activism, prior voting, and party evaluations significantly influence support for the MPP and DP ($p < 0.05$). Age shows a negative association with MPP and DP support, whereas political activism positively predicts DP support, reflecting differences in voter behavior. Model fit indicators (Cox and Snell $R^2 = 0.745–0.791$, Nagelkerke $R^2 = 0.803–0.852$, McFadden $R^2 = 0.519–0.593$) and the LR χ^2 test ($p < 0.001$) confirm strong explanatory power.

Classification results indicate high predictive accuracy for MPP and DP choices, ranging from 94–98% for MPP and 84–91% for DP across the two districts. In contrast, prediction for the "Other" category is relatively weak, reflecting the small number of observations and higher heterogeneity. Overall prediction accuracy reaches 78.3% in the 2nd district and 86.3% in the 4th district, demonstrating that the model systematically and reliably forecasts major party choices.

In conclusion, these findings integrate the demographic composition, latent variable structure, and MLR model performance, confirming that voter behavior and party support for major parties are well captured. The relatively weak prediction for the "Other" category highlights the need for further investigation into the heterogeneity of minor parties and item-level internal consistency

REFERENCES

Bank, W. (2022). *Worldwide governance indicators and voter behavior analysis.* World Bank Publications.

Batbold, T., & Tumenjargal, D. (2018). Factors influencing voter participation in Mongolia. *Politology, 18*(2), 45-58.

Blais, A. (2000). *To vote or not to vote: The merits and limits of rational choice theory.*

Brito, K., & Adeodato, P. (2023). Machine learning for predicting elections in Latin America based on social media engagement and polls. *Government Information Quarterly, 40*(1), 101782. doi:10.1016/j.giq.2022.101782

Downs, A. (1957). *An economic theory of democracy.*

Gueuwou, S., & Mensah, R.-M. (2023). *Stance prediction and analysis of Twitter data: A case study of Ghana 2020 presidential elections.* arXiv. Retrieved from https://arxiv.org/abs/2306.14203

Hair, B., & Babin, A. (2019). *Multivariate data analysis.* Cengage.

Hosmer, L. (2013). *Applied logistic regression.* Wiley.

International Foundation for Electoral Systems. (2021). *Voter participation and public trust in elections.* IFES.

Myagmartsooj, N. (2024). Factors influencing voter turnout in Mongolia (Youth case study). *Politology, 22*(1), 111–118.

National Statistics Office of Mongolia. (2023). *Population and election participation statistics.* Ulaanbaatar.

Norris, P. (2014). *Why electoral integrity matters.* Cambridge University Press.

Sant Maral Foundation. (2022). *Public opinion survey on political participation in Mongolia.* Ulaanbaatar.

Shahi, G., Basyurt, A., & Stieglitz, S. (2025). Agenda formation and prediction of voting tendencies for European Parliament election using textual, social and network features. *Information Systems Frontiers*. doi:10.1007/s10796-024-10568-w

Urandelger, G., Burmaa, S., & Altan-Od, M. (2023). Factors influencing youth political choice in Mongolia. *Journal of Business and Innovation, 9*(2), 55–70.

Хавсралт

СЭДЭВ: СОНГУУЛИЙН СОНГОЛТЫГ УРЬДЧИЛСАН ТААМАГЛАХ СУДАЛГАА

Зорилго: Сонгогчдын үндсэн шинж болон улс төрийн үнэлгээнд тулгуурлан намын сонголтын магадлалыг тооцож, сонгуулийн үр дүнг урьдчилан таамаглах зорилго бүхий судалгаа юм. Судалгаанд оролцогчдын мэдээллийг зөвхөн эрдэм шинжилгээний ажил, өгүүлэл бичихэд ашиглах бөгөөд ямар нэгэн **улс төрийн сонирхолгүй** болно. Судалгаанд оролцогчдын мэдээллийг хуульд заасны дагуу хамгаалж, нууцлах болно.

Судалгаа өгөхийг зөвшөөрч байна. Судалгаанд оролцогчийн утасны дугаар:
.....

Иргэний үнэмлэх дэх албан ёсны дүүрэг, аймаг бичих:
.....

СУУРЬ МЭДЭЭЛЭЛ:

1. Таны нас: .....
2. Таны хүйс: ☐ Эрэгтэй ☐ Эмэгтэй
3. Таны боловсролын түвшин:

☐ Бага ☐ Дунд ☐ Бүрэн дунд ☐ Диплом ☐ Бакалавр ☐ Магистр ☐ Доктор

4. Танай өрхийн сарын дундаж орлогын хэмжээг сонгоно уу?

☐ 1 сая төгрөгөөс доош ☐ 1-2 сая ☐ 2-3 сая ☐ 3-5 сая ☐ 5 сая дээш

УЛС ТӨРИЙН ОРОЛЦОО, ХАНДЛАГА:

1. Та улс төрийн мэдээллийг ихэвчлэн хаанаас авдаг вэ?

☐ ТВ ☐ Facebook ☐ TikTok ☐ Сайтууд ☐ Танил, найз нөхөд ☐ Бусад

2. Та өөрийнхөө улс төрийн идэвхийг үнэлнэ үү?

☐ Маш идэвхгүй ☐ Идэвхгүй ☐ Дунд ☐ Идэвхтэй ☐ Маш идэвхтэй

3. Та өмнөх УИХ сонгуульд саналаа өгсөн үү?

☐ Тийм ☐ Үгүй ☐ Санахгүй байна

4. Та дараах улс төрийн намуудын бодлого, үйл ажиллагааг нийтэд нь хэрхэн үнэлж байна вэ?

Намын нэр	**Маш муу**	**Муу**	**Дунд**	**Сайн**	**Маш сайн**
МАН					
АН					
ХҮН нам					
Бусад нам					

5. Хэрэв өнөөдөр сонгууль болбол та аль намд саналаа өгөх вэ? (Заавал хариулна.)

☐ МАН ☐ АН ☐ ХҮН нам ☐ Бусад нам