The Role of Social Media Platforms in Forecasting Elections: A Comparison of Twitter and Facebook

TAPIO VEPSÄLÄINEN, Information Systems Science, Department of Management and Entrepreneurship, Turku School of Economics, University of Turku, Turku, Finland
HONGXIU LI, Department of Information and Knowledge Management, Faculty of Management and Business, Tampere University, Tampere, Finland
REIMA SUOMI, Information Systems Science, Department of Management and Entrepreneurship, Turku School of Economics, University of Turku, Turku, Finland

Prior literature shows that social media could be used to forecast political elections. Most studies have focused on a single social media platform, and few studies have explored the use of social media data across multiple platforms to make election predictions. Though candidates' personal attributes have also been suggested as critical factors affecting election results, there has been little research into the interacting effect of social media and candidate attributes in predicting elections. To address the research gap, this article investigates the role of two different social media platforms, Twitter (now known as "X") and Facebook, in forecasting the 2019 Finnish parliamentary elections and how candidates' political experience moderates the role of the two platforms in predicting elections. The findings show that both the number of Facebook likes and the number of Twitter followers are associated with the election outcome positively. Political experience of candidates moderates the association between the number of Facebook likes and election outcomes as well as the association between the number of Twitter followers and election outcomes. This research adds to the discussion on how social media can predict election results. It considers both different social media platforms and the roles of the candidate attributes.

CCS Concepts: • **Information systems** → **Web mining**;

Additional Key Words and Phrases: Social media, Facebook, Twitter, elections, forecasting, political experience

ACM Reference Format:
Tapio Vepsäläinen, Hongxiu Li, and Reima Suomi. 2024. The Role of Social Media Platforms in Forecasting Elections: A Comparison of Twitter and Facebook. *Digit. Gov. Res. Pract.* 5, 3, Article 19 (September 2024), 15 pages. https://doi.org/10.1145/3651227

1 INTRODUCTION

Social media has transformed the production, distribution, and consumption of political information [1]. As a result, platforms such as Facebook and Twitter have become an integral part of political campaigning.

Authors' addresses: T. Vepsäläinen (Corresponding author) and R. Suomi, Information Systems Science, Department of Management and Entrepreneurship, Turku School of Economics, University of Turku, Turku, Finland; e-mail: Tapio.a.vepsalainen@utu.fi; H. Li, Department of Information and Knowledge Management, Faculty of Management and Business, Tampere University, Tampere, Finland.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM 2639-0175/2024/09-ART19
https://doi.org/10.1145/3651227

Empirical evidence suggests that political campaigns combining social media with traditional methods are more successful than those relying solely on conventional strategies [2, 3]. The importance of social media has been steadily growing in recent years, and politicians are increasingly using social media to inform and communicate with their followers [2]. For instance, Facebook adoption among political candidates has diffused rapidly in the US since 2006 [4]. By 2008, 72% of major party House of Representatives candidates had already set up their Facebook pages for political campaigning [4]. Twitter appears to become another social media integrated into political communication [5]. Prior research has reported on the application of social media in elections in various countries and the use of different social media in political candidates' campaigns [6–8].

The application of social media by candidates has led to an abundance of data about citizens' preferences or sentiments about candidates. This data has opened a new realm of possibilities for researchers to predict elections via the use of social media data, which might provide a better understanding of the electorate and their potential voting behavior. Using social media data to forecast elections has its potential benefits over traditional methods. The analysis of social media conversations and trends may predict the outcome of an election in real time [9–11]. Additionally, using social media data to forecast elections may require smaller investments than opinion polls, making it an attractive option for both private and public organizations [12–14].

Although researchers have explored the potential of using social media data to predict election results, there is still no consensus on how effective this approach can be [15–19]. So far, empirical evidence suggests that social media is associated with the outcome of elections. For instance, Twitter followers have been argued to be an indicator to understand the political preference of a candidate in a population [17, 18, 20]. Similarly, separate research has identified a correlation between the number of Facebook likes and election outcomes [4, 21]. However, variations in the sampling methods of social media data have been shown to yield inconsistent results [22]. Subsequent research indicates that the initial reports of high accuracy in using Twitter to forecast elections may have been overstated [23] and that the sentiment of tweets may not be as clearly associated with voting intention as expected [24].

Though political candidates have used different social media for their campaigns, prior studies have focused mainly on a single social media platform [1]. Only few studies have explored the use of cross-platform social media data to make predictions, which could provide an understanding of the roles of different social media in predicting elections [17, 25]. Recent findings have indicated that there is a varying correlation between the attention garnered on social media and subsequent election outcomes across countries and platforms [25]. Thus, there is a need for research to compare the role of different social media platforms in elections to gain a better understanding of how social media could be used to effectively predict election outcomes.

Candidates' personal attributes have also been found to be associated with election outcomes [26]. However, prior research has scarcely considered integrating social media data and candidates' personal attributes to predict election outcomes, and there is a lack of knowledge of how candidates' personal attributes could moderate the role of social media in elections. Thus, it is meaningful to examine the interaction role of candidates' personal attributes and social media in elections, which could provide an understanding of election prediction from the joint view of both social media and candidates' personal attributes.

Despite a growing body of research on the influence of social media on election outcomes, studies often consider only a single platform, particularly Twitter, while the combined effect of multiple platforms is overlooked. Furthermore, these analyses commonly disregard the possible moderating role of candidates' personal attributes. To address this gap, this study answers the following two research questions: (i) what is the role of social media platforms, such as Twitter and Facebook, in predicting election outcomes? And (ii) How do candidates' personal attributes moderate the role of social media platforms, such as Twitter and Facebook, in predicting election outcomes?

Building upon the research question, we compare the role of Twitter followers and Facebook likes in predicting the outcome of the Finnish Parliament elections in 2019 and the moderating effect of candidates' political experience. This approach might position our study among the first to compare the impact of different social

media platforms in forecasting elections and the moderating effect of candidates' personal attributes. Thus, this study expands the knowledge base by presenting a comparative analysis of the role of Facebook and Twitter activity on election outcomes and addressing how candidates' political experience influences the relationship between social media activity and election outcomes.

The remainder of the article is structured as follows: In Section 2, a literature review on social media in elections and the context for the study is presented. Section 3 introduces the research method applied in this study. The research results and the research findings are discussed in Section 4. Finally, we present our conclusions based on the research findings.

2 LITERATURE REVIEW

2.1 Research on Social Media in Elections

Extant literature has researched on the impacts of social media on elections, and social media has been found to have an impact on voting behavior in elections. The most widely studied example is the 2016 US Presidential Elections, where both Donald Trump and Hillary Clinton invested heavily in social media to reach out to voters and promote their respective campaigns [27]. Trump's use of the platform has been particularly controversial, with allegations of false information being spread to sway voters [28]. Some scholars have validated the effects of social media on elections in the 2017 Italian general election [29], the 2017 Ecuadorian presidential elections [30], 2018 Brazilian presidential elections [31], and the 2018 Malaysian general election [32].

Twitter has been the most studied social media in elections [15–19, 33]. Prior studies have examined the role of Twitter in elections, mainly analyzing the count of followers and the sentiment of the text messages sent by users on Twitter. A couple of studies have investigated the relationship between candidates' popularity on Facebook and election results and the results have been promising. For instance, Zhang [8] found that the popularity of candidates' Facebook posts is positively related to their election results. MacWilliams [34] argued that when combined with standard forecasting fundamentals, Facebook likes can produce accurate forecasts. Vepsäläinen et al. [21] reported that Facebook likes are a significant indicator of electoral success. Chang, Chiang [35] found that an ideology measure and Facebook support rate can predict actual election results. Evidence from Denmark suggests that Facebook likes are especially useful for predicting voting intentions in a multi-party setting [36]. These studies suggest that there is a positive link between Facebook campaigning and electoral success.

Some studies have also found differing results [37–39]. Sandoval-Almazan and Valle-Cruz [37] demonstrated that while Facebook reactions can increase voter engagement, they are not correlated with electoral outcomes. Social media engagement may not always represent genuine support that translates to votes in elections [38]. Some users engage with social media content for entertainment, or, for staying informed about political opinions and policies, or out of curiosity without necessarily agreeing with the politician or any intention of voting for the candidate. Moreover, social media metrics can be influenced by bots or paid campaigns, artificially inflating the perceived popularity of a candidate [39]. Additionally, followers on social media may not be eligible voters due to age, nationality, or other disqualifying factors. Individuals' voting behavior is complex and not solely determined by social media consumption [40].

A growing number of empirical research articles have been published in the field of social media in elections, and some scholars have reviewed literature to provide a full picture of research on social media in elections and to identify areas requiring further investigation (see Table 1). These studies conclude that there are many challenges associated with predicting election outcomes with social media, though some promising results have been made [15–19, 33, 41]. One identified research gap is the relatively high focus on Twitter and less attention to other social media platforms, such as Facebook. Twitter has a relatively small user base, and other larger social media platforms such as Facebook could offer better opportunities for retrieving a representative sample. However, prior studies have not paid much attention to exploring data from other social media. In addition, little research has attempted to combine social media data from different social media platforms, such as Facebook

Table 1. Review of Relevant Research

Authors	Summary from the literature review
Rousidis et al. [15]	The results of the study are inconclusive, with only 53.1% of the studies managing to achieve a valid prediction. Furthermore, the majority of researchers (over 75%) use data from Twitter only, which may not be enough to generate highly accurate predictions.
Gayo-Avello [16]	Previous research on the predictive power of Twitter has shown that it is often overestimated. It is recommended to use predicted vote rates rather than winner prediction and number of correctly guessed races, as the latter can be due to chance. Additionally, simple baselines have the potential to achieve better accuracy than Twitter predictions in many cases.
Brito et al. [17]	Overall, most studies examining the connection between online sentiment and election outcomes have been conducted in the context of a single election and using Twitter data. Studies relying solely on Twitter data have demonstrated significantly lower success rates than those based on other social networks, such as Facebook.
Chauhan et al. [18]	Most studies on election predictions have used Twitter as the corpus, with the majority of them being successful in their predictions. This demonstrates the value of Twitter in understanding public opinion and accurately predicting election outcomes.
Phillips et al. [19]	Existing research relies on relatively simple methods, ranging from linear regression to keyword matching, but results that are successful in one context often fail in others. By incorporating domain-specific knowledge, researchers can help point their statistical models in the right direction to produce better results.
Santos et al. [33]	Social media analyses may provide valuable insights into public opinion, but they are not a substitute for traditional polls. Most published studies focus on count-based approach based on Twitter data. The forecasting accuracy of studies was deemed ambiguous. Further research is required to fully understand the potential of social media analyses for public opinion polling. Additionally, the sharing of datasets and information among researchers is essential for advancing the field.
Skoric et al. [41]	This study has shown that while combining multiple sources of data can improve prediction performance, there are not enough studies using this approach to draw definitive conclusions. Moreover, social media analytics should not replace survey-based studies, but rather should be used to gain additional insights into public opinion and political behavior. Last, there is a need for more theoretically informed work in the field that pays attention to the underlying mechanisms and processes.

and Twitter, in making election predictions and helping understand the different roles of social media in these predictions. There is a need for studies exploring other data sources, especially those that combine multiple social media platforms [15, 17].

2.2 Research on Candidates' Personal Attributes in Elections

Extant research has studied the role of **personal vote-earning attributes (PVEAs)** in elections, such as political experience, gender, and occupational background [42]. The role of candidates' attributes is especially relevant under open-list proportional representation due to the individualistic dimension of competition [43]. The PVEAs act as information cues to voters who are facing difficulties in selecting their preferred candidate [43].

PVEAs can be seen as personal traits that demonstrate candidates' credibility and aptitude for representing various voter subgroups [26]. Candidate involvement in local and party politics, for example, may signal

their dedication to the public and their ability to gain their trust, which might be linked with the candidates' success in elections [26]. Gender, celebrity status, occupational background, and especially political experience have been suggested to have impacts on voting decisions [44–46]. Prior literature also found that differences between districts and parties could lead to disparate behavior [2]. Some scholars have argued that candidates' campaign spending could improve electoral success [47]. The literature also found that intra-party competition could increase candidates' appeal to the electorate by focusing on specific themes, such as their position as local candidates to distinguish themselves from their fellow party members [48]. Moreover, the position of candidates could also be associated with election success. Prior research shows that electoral experience at the municipal level and district nativity is more important to rookie candidates than to long-term politicians [48].

Prior studies have mainly focused on examining the link between PVEAs and election outcomes, few studies have considered the interactive role of PVEAs and social media, though both dimensions have been identified to be closely linked with election success. There is a need for research to understand elections from the view of both social media and candidates' personal attributes. Thus, this study will examine the moderating role of candidates' personal attributes, specifically political experience, between social media and election outcomes.

2.3 The Case for Finnish Elections

Research on election forecasting based on social media has been conducted in at least 29 different countries [17]. The electoral systems and cultures of each country have their own distinct features. Most studies have focused on elections with relatively few candidates, such as the US presidential elections [15–18]. However, research suggests that the number of social media followers is a particularly useful metric for discerning voting intentions in multi-party systems, like those found in Finland [36].

The Finnish political system is an open-list proportional representation system with multiple parties. Citizens select their preferred candidate from a large list of contenders. The votes are pooled to the party, and the total number of preference votes determines the number of seats won. Electoral competition is not limited to competing against other parties, as candidates are also competing with their fellow party members. Intraparty competition encourages candidates to focus on their individual campaigns and differentiate themselves from fellow party members while adhering to party policies [43].

The electorate faces challenges in deciding their votes. Each citizen must select a candidate from a long list of contenders, and as a result, voters must digest a substantial amount of relevant information to support their decisions on votes. The electorates are likely to resort to heuristics and cognitive short-cuts when casting their vote [42, 45].

Academic studies on forecasting Finnish elections are scarce. Forecasts presented in the media are largely based on opinion polling [49]. There are three prominent opinion research institutions that are the main providers of opinion polling in Finland [49]. Other methods, such as statistical and econometric models, are much less popular in forecasting Finnish elections. Vepsäläinen et al. [21] examined the link between social media (Facebook) and election success and reported a positive and significant relationship between Facebook likes and electoral success. Clearly, more research is needed to help understand the role of social media in predicting elections in Finland, which could provide evidence to election candidates and their parties on the value of social media in elections and how to best utilize it.

3 RESEARCH METHOD

To answer the research questions, we first collected social media data among Finnish politicians from Facebook and Twitter as well as the demographic and political data about the political candidates from Finland. Then, we combined these different data and cleaned the dataset to get complete data for the analysis. After data cleaning, we have performed data transformation to make the dataset fit to the multiple linear regression analysis using the **Ordinary Least Squares (OLS)**.

3.1 Data

The 2019 Finnish parliamentary elections were held on the 14th of April, with a total of 2,468 candidates nominated. Social media data was gathered on the 4th of April, 2019, from the two most popular platforms among Finnish politicians: Twitter and Facebook. The list of candidate pages was manually compiled and updated. Facebook data was collected from candidates' public profiles without logging in to Facebook, resulting in data for 1,502 candidates. Twitter data, specifically follower counts, was collected using the Twitter API, and accounts were found for 1,466 candidates.

Social media data was collected with a 10-day lead time before the election. This approach was based on the understanding that the number of Facebook likes and Twitter followers are relatively static figures, not subjected to the rapid fluctuations that characterize other social media metrics such as sentiment or engagement levels. Consequently, we did not anticipate that the lead time would significantly impact our findings, thus maintaining methodological robustness for the static metrics evaluated within this study's timeline.

A list of candidates and additional relevant details was compiled from Statistics Finland. This dataset included a range of demographic and political information, such as sex, age, occupation, and party affiliation. It also encompassed candidates' electoral district, incumbency status, past membership in the European Parliament, previous service in the Finnish parliament, participation in the last parliamentary elections, and any history of election to municipal government positions [50].

The details related to the self-reported campaign budget and education were collected using web scraping from the national broadcasting company's (Yle) voting advice applications [51]. Yle hosts a popular voting compass service, free for any citizen to use, where information about relevant candidates can be found. Each candidate fills in their information to the service themselves.

The statistical method selected for this study relies on the assumption that the dataset is complete. To ensure accurate results, it is important to remove entries that are unlikely to be relevant to the research questions in this study. In the initial data, a significant portion of candidates utilized social media platforms, with 75% present on Facebook and 67% on Twitter. Meanwhile, a small minority of 15% were not present on either Facebook or Twitter. A majority of candidates were transparent about their campaign finances and educational background, with 87% providing information related to their campaign budget and 85% disclosing their education level. The final sample was cleaned to only include candidates who were using both Facebook and Twitter and had reported their personal information, such as age, party, district, occupation, campaign budget, education, and sex. There were 1,022 candidates left after this round of cleaning.

The sample was further reduced to only include candidates presenting parties that had realistic chances to have their candidates elected. The removal was done to reduce the number of categories for the party variable, as there are many parties with only a few candidates. The candidates whose party had not had any seats in the parliament in the past, whose party received total votes less than 25,000 in the 2015 elections, and whose party has less than 5,000 Facebook likes in total were removed from the sample. This resulted in the removal of 74 candidates, leaving a final sample of 948 candidates.

3.2 Variables

Election success is the **dependent variable (DV)** in this study, which was measured by the count of preference votes for a candidate in the election. The **independent variables (IV)** of the study are the number of Twitter followers and Facebook likes. Facebook likes and Twitter followers were selected to ensure simplicity, as earlier studies have reported a high correlation level among various measures of online engagement [21].

Following the prior literature on PVEAs, since candidates' personal attributes are linked with election outcomes, it is reasonable to set some candidates' personal attributes as control variables in this study. Specifically, seven variables about the candidates and their party are set as control variables in this study, including age, party, district, occupation, self-reported campaign budget, education level, and sex. Political experience is set

Table 2. Variables and Measurements

Variable		Measurement
Dependent variable	Election success	The number of preference votes for a candidate in an election
Independent variable	Twitter followers	The number of Twitter followers of a candidate
	Facebook likes	The number of Facebook likes of a candidate
Moderator	Political experience	A sum of incumbency, status as a member of the European Parliament, whether they have been in the parliament in the past, whether they ran for the last parliament elections, and if they hold the municipal office. Each included variable is represented as 0-1 (0: no experience, 1: have experience) in the initial data. The final value for political experience is the sum of those values. (0-5, ordinal variable, higher value means more experience)
Control variable	Age	The age of a candidate
	Party	Which party a candidate sits in (10 parties: PS, KOK, KESK, KD, VIHR, SIN, LN, VAS, SDP, RKP)
	District	District of candidacy (13 districts: HEL, UUS, VAR, SAT, AHV, HÄM, PIR, KAA, SKA, VAA, KES, OUL, LAP)
	Occupation	Occupation of each candidate (Members of Parliament, Specialists, Managers, Experts, Entrepreneurs, Construction, Repair and Manufacturing Workers, Service and Sales Workers, Students, Office and Customer Service Workers, Farmers, Forestry Workers, and so on, Other Employees, Soldiers, Pensioners, Process and Transportation Workers, Unemployed, Commissioners)
	Campaign budget	Campaign budget category self-reported by candidate (1-5 code) (1: < 5,000 EUR; 2: 5,000–10,000 EUR; 3: 10,000–20,000 EUR; 4: 20,000–50,000 EUR; and 5: > 50,000 EUR)
	Education	Education level self-reported by candidate (1: primary education; 2: vocational education; 3: high school education; and 4: higher education)
	Sex	Female or male (0: male; 1: female)

Note: PS: True Finns, KOK: National Coalition, KESK: Centre party, KD: Christian Democrats, VIHR: The Green party, SIN: Blue Reform, LN: Movement Now, VAS: Left Alliance, SDP: Social Democratic Party, RKP: Swedish People's party
HEL: Helsinki, UUS: Uusimaa, VAR: Finland Proper, SAT: Satakunta, AHV: Åland, HÄM: Tavastia, PIR: Pirkanmaa, KAA: South-East Finland, SKA: Savonia-Karelia, VAA: Vaasa, KES: Central Finland, OUL: Oulu, LAP: Lapland)

as a moderator. We simplified the model by combining known political positions into a single variable called political experience [42]. It is a combination of incumbency, status as a member of the European Parliament, whether they have been in the parliament in the past, whether they ran for the last parliament elections, and if they hold the municipal office.

Some variables were transformed into a more appropriate form for analysis to keep the model as simple as possible. Details of these variables included in this study are included in Table 2.

3.3 Data Analysis

Multiple linear regression using the OLS was selected for model estimation. The suitability of OLS is confirmed through checking for linearity, independence, homoscedasticity, normality, and multicollinearity [52].

The multiple linear regression model's evaluation criteria for election outcomes are: (1) establishing statistical significance with p-values less than 0.05; (2) assessing the strength of predictions through substantial beta coefficients that indicate practical significance; and (3) gauging the model's explanatory power by a high adjusted R squared value, which reflects the proportion of variance in election outcomes that the model accounts for.

Table 3. Descriptive Statistics

Variable	Mean	Std	Min	Max	Kurtosis	Skewness
Votes	2,167.43	2,798.21	22.00	24,542.00	13.29	3.07
Facebook Likes	1,327.50	3,161.45	3.00	51,663.00	97.51	8.35
Twitter Followers	2,004.77	9,347.57	1.00	131,679.00	134.52	10.87
Age	42.94	11.82	18.00	76.00	−0.60	0.14
Campaign budget	2.46	1.15	1.00	5.00	−1.14	0.17
Education	3.51	0.81	1.00	4.00	0.34	−1.35
Political experience	0.99	1.14	0.00	4.00	1.97	1.62

Table 4. Correlation Test Results

	PV	FL	TF	AG	CB	ED	PE	OC	PA	DI
Preference votes (PV)	1.0									
Facebook likes (FL)	0.50***	1.0								
Twitter followers (TF)	0.38***	0.51***	1.0							
Age (AG)	0.03	−0.13***	−0.12***	1.0						
Campaign budget (CB)	0.55***	0.44***	0.37***	0.04	1.0					
Education (ED)	0.12***	0.08**	0.15***	0.04	0.16***	1.0				
Political Experience (PE)	0.53***	0.36***	0.31***	0.12***	0.45***	0.07*	1.0			
Occupation category (OC)	0.31***	0.23***	0.28***	0.17***	0.30***	0.34***	0.34***	1.0		
Party (PA)	0.26***	0.13***	0.05	0.02	0.22***	−0.01	0.15***	0.02	1.0	
District (DI)	0.07**	0.02	0.03	−0.03	0.07**	−0.02	0.05	0.02	−0.01	1.0

Note: *p<0.05, **p<0.01, ***p<0.001.

Table 3 presents the summary statistics of the variables used to assess the appropriateness of OLS regression. The distributions of vote counts, Facebook likes, and Twitter followers are positively skewed, as indicated by the skewness and kurtosis values.

The count of Facebook likes and the number of Twitter followers both have a strong and significant correlation with the number of votes received by the candidate. Kendall's τ coefficient for Facebook likes is 0.50 ($p < 0.001$) and Twitter followers are 0.38 ($p < 0.001$). Campaign budget scale and political experience both have a higher correlation. The subsequent coefficients were 0.55 ($p < 0.001$) for campaign budget scale and 0.53 ($p < 0.001$) for political experience. A summary of the correlations is in Table 4.

A logarithmic transformation was conducted on the number of preference votes, Facebook likes, and Twitter followers because the data was skewed (see Table 3). This rescaling helps with data normalization when the data is right-skewed or has a wide range of values, which is a requirement for OLS in regression analysis.

Figure 1 provides visual cues for inspecting relationships between the dependent and independent variables after the logarithmic conversion. The scatter plots and trend lines suggest a positive coefficient between the independent and dependent variables. The plots have been colored according to whether a candidate was elected. The scatter plots for Facebook likes and Twitter followers seem to be overlapping. Their internal Kendall's τ correlation coefficient was inspected at 0.50 ($p < 0.005$), suggesting that there may be multicollinearity issue.

We have addressed the possible multicollinearity by calculating the **variance inflation factors (VIF)** for the variables included in this study. The highest VIF is 2.15, suggesting multicollinearity does not pose a concern in the OLS regression analysis.

After the logarithmic conversion, the variables were standardized using z-score standardization, which involves removing the mean from each feature and scaling it to unit variance using the scikit-learn Standard

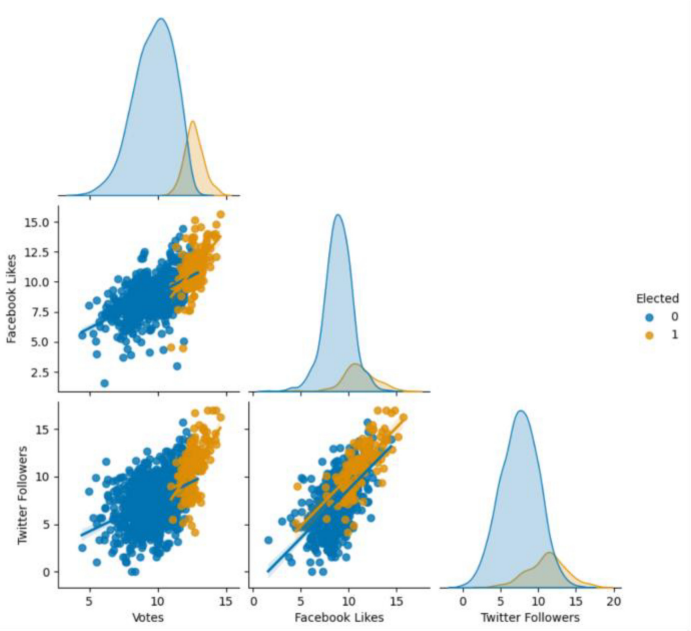

Fig. 1. Visual inspection of the relationship between IVs and DV.

Scaler [53]. Z-score standardization is performed here, because the focus is on comparison of the platforms. It neutralizes scale discrepancies among features, allowing for a fair comparison and interpretation of coefficients [54].

After ensuring that the data meets the criteria for OLS regression, we developed multiple models to analyze the impact of social media presence and political experience on election outcomes. This step-by-step approach allows for a detailed analysis of how each factor, alone and in concert with others, can impact the dependent variable of votes received.

Following model analysis, cross-validation is performed to confirm the reliability of findings. By dividing the data into subsets, training on one and validating on another, cross-validation reduces the risk of overfitting and ensures the model's predictive capabilities across different data subsets [52].

The OLS method was applied in data analysis using Python 3.10 and statsmodels v0.13.2 [55]. The model cross-validation was done with scikit-learn [53]. We also explored the relative importance of the control and independent variables using Pingouin 0.5.2. The relative importance allows us to compare how different variables contribute to the model [56].

4 RESULTS

A series of regressions was conducted to test whether Facebook likes and Twitter followers are associated with election outcomes (votes) and the moderating role of political experience on their relationship (see Table 5). For each model, we checked the VIF value. The highest VIF value in the overall models was 2.74, which was below the vigilance threshold of 5.0 [57], indicating that multicollinearity was not a serious issue for each regression model.

Model 1 is the control model with district, party, campaign budget, sex, age, occupation category, and education level as control variables. District, party, campaign budget, and sex were found to be significantly associated with election outcomes. This indicates that these variables should be taken into consideration when making election

Table 5. Regression Analysis Results

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Intercept	8.48*** (0.18)	7.92*** (0.18)	8.47*** (0.17)	8.53*** (0.17)	8.35*** (0.18)	8.42*** (0.18)	8.63*** (0.17)
Age (A)	−0.06 (0.04)	−0.05 (0.04)	0.10** (0.03)	0.09** (0.03)	0.04 (0.04)	0.04 (0.04)	0.10** (0.03)
Campaign budget (CB)	0.75*** (0.04)	0.64*** (0.04)	0.45*** (0.04)	0.44*** (0.04)	0.55*** (0.04)	0.53*** (0.04)	0.42*** (0.04)
Education (E)	−0.05 (0.05)	0.03 (0.04)	0.04 (0.04)	0.03 (0.04)	−0.01 (0.04)	−0.02 (0.04)	0.02 (0.04)
Sex (S)	0.19*** (0.04)	0.17*** (0.04)	0.14*** (0.03)	0.13*** (0.03)	0.17*** (0.03)	0.16*** (0.03)	0.13*** (0.03)
District (D)	0.07 (0.04)	0.06 (0.03)	0.07* (0.03)	0.07* (0.03)	0.06 (0.03)	0.06 (0.03)	0.07* (0.03)
Party (P)	0.37*** (0.04)	0.37*** (0.04)	0.35*** (0.03)	0.33*** (0.03)	0.39*** (0.03)	0.37*** (0.03)	0.33*** (0.03)
Occupation category (OC)	0.52*** (0.04)	0.05 (0.06)	−0.03 (0.06)	0.03 (0.06)	−0.04 (0.06)	0.04 (0.06)	0.06 (0.06)
Political Experience (PE)		0.55*** (0.05)	0.44*** (0.05)	0.48*** (0.05)	0.49*** (0.05)	0.55*** (0.05)	0.51*** (0.05)
Facebook likes (FL)			0.57*** (0.04)	0.68*** (0.05)			0.56*** (0.06)
Moderator: FL * PE				−0.10*** (0.03)			−0.02 (0.04)
Twitter followers (TF)					0.36*** (0.04)	0.46*** (0.05)	0.20*** (0.06)
Moderator: TF * PE						−0.11*** (0.03)	−0.13** (0.05)
Adjusted R squared	0.590	0.631	0.694	0.698	0.655	0.659	0.701

Note: Standard errors are in parentheses. ***p < 0.001, **p < 0.01, *p < 0.05.

forecasting. Similarly, Model 2 suggests that candidates' political experience is also positively associated with election outcomes.

We then tested the link between Facebook likes and election outcomes (Model 3). The results suggest that the numbers of Facebook likes are positively linked with election outcomes. Next, we tested the moderating effect of political experience on the relationship between Facebook likes and election outcomes (see Model 4). As shown in Model 4, there is a negative and significant moderating effect of political experience on the relationship between Facebook likes and election outcomes ($\beta = -0.10$; p <0.001). This suggests that political experience moderates the relationship between Facebook likes and preference votes. Specifically, when candidates have more political experience, the relationship between Facebook likes and preference votes becomes weaker.

We then tested the link between Twitter followers and election outcomes. As shown in Model 5, the number of Twitter followers has a positive and significant link with election outcomes ($\beta = 0.36$, $p < 0{,}001$). This means that the number of Twitter followers of a candidate is positively linked with election outcome (votes). We also tested the moderating effect of political experience on the relationship between Twitter followers and election outcomes (see Model 6). The results indicate that there is a negative and statistically significant moderating effect of political experience on the relationship between Twitter followers and election outcomes ($\beta = -0.11$; p <0.001). The test results indicate that political experience moderates the relationship between Twitter followers and preference votes negatively. When candidates have more political experience, the link between Twitter followers and election outcomes (votes) will be weaker.

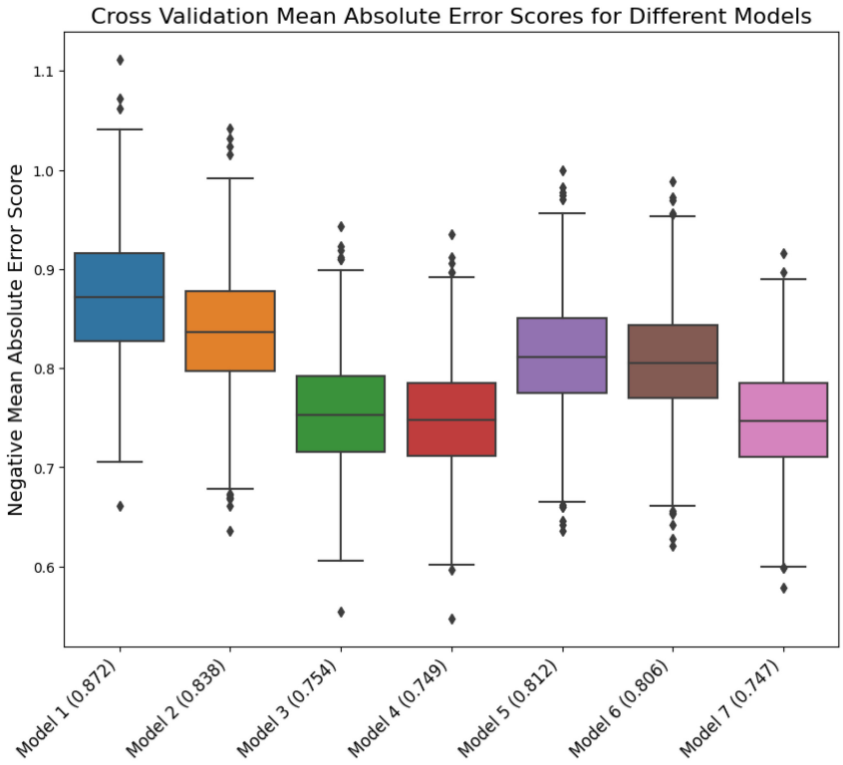

Fig. 2. K-fold cross-validation of different models.

When we include both Twitter followers and Facebook likes in the regression model (see Model 7), the relationships between Facebook likes ($\beta = 0.56$, $p < 0.001$) and election outcomes as well as between Twitter followers ($\beta = 0.20$, $p < 0.001$) and election outcomes are positive and significant. Political experience keeps its negative moderating effect on the relationship between Twitter followers and election outcomes ($\beta = -0.13$, $p < 0.001$) but loses its moderating effects on the relationship between Facebook likes and election outcomes.

One way to measure the predictive capability of the models is to look at the adjusted R2. Model 7 had the best performance, with an R2 of 0.701. Comparing different metrics only by coefficients or R2 is not enough to answer our research questions [56]. R-squared statistic can be used to measure how well a model fits the data and how much of the variation in the data is explained by the model, but it does not ensure that the model would work with unseen data. We therefore further validated the results by using a k-fold cross-validation.

The algorithm divides the data into k-test and training sets and repeats the procedure n times. To keep it simple, we use k = 10 splits with n = 100 repeats. The comparison between models is based on the **Mean Absolute Error (MAE)**, where a lower MAE indicates a model that predicts more closely to the observed number of votes. It is important to consider that due to the transformation of the data, the interpretation of the MAE may be influenced, and it should be used primarily for comparative purposes between models.

The cross-validation indicates that incorporating Facebook likes into the model reduces the MAE, suggesting a comparative improvement over the control model. Also, adding the interaction of Facebook likes and political experience decreases the MAE slightly. Adding Twitter term does not reduce the MAE. The MAE for the control model is 0.838. Facebook model with control variables has MAE of 0,754. The model with both Twitter followers and Facebook likes also has MAE of 0,754. The lowest MAE is for Model 8 (0,747). Figure 2 illustrates these comparative MAE values across the models.

Table 6. Relative Feature Importance

Variables	Relative importance (R^2 partitioned by averaging over orders)	Relative importance (percentage)
Campaign budget (CB)	0.15	20.6%
Facebook likes (FL)	0.13	18.07%
Political Experience (PE)	0.08	11.63%
Moderator: FL * PE	0.08	11.45%
Moderator: TF * PE	0.07	10.11%
Party (PA)	0.07	9.95%
Twitter followers (TF)	0.07	9.3%
Occupation category (OC)	0.05	6.56%
Sex (SE)	0.00	0.7%
Education (ED)	0.00	0.66%
Age (AG)	0.00	0.51%
District (DI)	0.00	0.48%

The test results suggest that Facebook likes have a stronger association with election outcomes compared with Twitter followers. Though Twitter followers could be a relevant variable in predicting election outcomes, their relevance in predicting election outcomes is weak, since both the R2 and the cross-validation suggest that Twitter followers could not add much value to the statistical models. To further understand the relevance of Twitter followers in predicting elections, we also calculated the independent variable importance metrics with the method introduced by Grömping [56]. Specifically, we applied the recommended metric lmg, which decomposes explained variance into non-negative contributions [56]. The relative importance is only calculated for the best-performing model (Model 7).

The relative importance metric suggests that the campaign budget is the most important metric (20.6%). Interestingly, Facebook likes are the second most important metric with a relative importance of 18.07%. Political experience is the third most important variable. While Twitter seems less relevant in the earlier analyses, in the relative importance analysis it does seem more relevant than occupation, district, sex, age, and education. The moderating terms have both higher relative importance than Twitter followers itself. The results are presented in Table 6.

5 DISCUSSION AND CONCLUSIONS

5.1 Discussion

This study compared the effects of social media platforms, specifically Twitter and Facebook, on predicting election outcomes and the moderating effect of candidates' personal attributes on their effects in Finnish parliamentary elections.

Our research found that Facebook likes and Twitter followers are positively associated with election outcomes, and Facebook likes have a stronger association with election success than Twitter followers. Candidates' political experience moderates the associations between Facebook likes and Twitter followers and election success. A model combining both social media data, candidates' political experience, and the interacting effects of these variables had the best performance in predicting election success. Ultimately, these findings suggest that understanding citizens' voting behavior and existing forecasting frameworks should incorporate both social media data and candidates' political experience in the forecasting models.

Our findings suggest that candidates with more Facebook likes and Twitter followers would get more votes from electorates. These findings are consistent with the findings in prior studies that have studied the role of social media in elections, including Facebook likes and Twitter followers. MacWilliams [34], for example, claimed

that Facebook likes are positively associated with election outcomes in the US, whereas Vepsäläinen et al. [21] concluded that Facebook likes are positively linked with election success in Finland.

The results of our tests revealed a statistically significant negative interaction between political experience and social media followers count (Facebook likes and Twitter followers). We found that the numbers of Twitter followers and Facebook likes are more influential in determining the election outcome for candidates with less political experience. This may suggest that in the digital era social media is a more efficient communication channel for candidates with less political experience, allowing them to reach a larger audience and potentially gain success, whereas for veteran politicians, they already have media coverage and supporters; social media might have weaker effect on election success compared to those candidates with less political experience.

5.2 Contributions

The research contributes to the existing knowledge on forecasting elections with social media data in a few ways. First, this study enriches the literature on social media in election prediction by taking both social media and candidates' personal attributes and their interaction effects into consideration in prediction. Our research demonstrates that combining social media data and candidates' personal attributes in predicting election outcomes could provide a better understanding of the potential of social media in forecasting elections. Including political experience in the prediction of election success in conjunction with Facebook and Twitter data seems to be more efficient than relying on only these social media data.

Second, this study indicates that the role of social media on election prediction differs between social media platforms. Our study compares the different roles of Facebook likes and Twitter followers and demonstrated that Facebook likes had a stronger relationship with election success than Twitter followers in our sample.

This study also presents the evidence of how social media's impact on predicting election success is moderated by candidates' political experience. The findings suggest political experience serves as a negative moderator, yielding insights into the interconnected effects of social media and candidates' characteristics.

Finally, this study offers some potentially practical insights on the relationship between social media use and electoral outcomes. This study examines which social media metrics might correlate with voting behaviors, which can inform campaigners how to utilize different social media platforms in their campaigns. The findings on the negative moderating effect of political experience on the relationship between social media and election success suggest that candidates without political experience might utilize social media as an effective tool in their campaigns.

5.3 Limitations

Our study has its limitations. Using social media data for election prediction faces some issues that should be taken into consideration. Our study focuses on numerical metrics and misses the context of user interactions. There might be fake accounts, bots, and other forms of manipulation in social media data that could misrepresent a candidate's popularity. It is also evident that social media users are unlikely to accurately represent the voter population, which affects the relationship between online metrics and actual votes. Moreover, the rapidly evolving nature of social media platforms, coupled with potential technical and privacy constraints, poses challenges for data collection and accuracy.

There are also some study-specific issues that should be considered in future research. First, the research focuses on Finnish elections, which limits how well the findings can be generalized to different countries with varying political systems. Future research should include comparisons with countries in other regions to assess the global applicability of the findings. Second, this study focuses on certain features and specific social media platforms, such as Twitter and Facebook; future research could consider including more social media platforms and different social media metrics in examining the role of social media in elections. Third, the data collection period may influence the results to a certain degree, which should be taken into consideration.

All these factors highlight the complexity of using social media data for election prediction and underscore the need for ongoing revision and enhancement of research methodologies in this field. It must be stressed that to verify the consistency of social media as a predictive tool, examining trends over multiple elections is necessary.

REFERENCES

[1] A. Casero-Ripollés. 2018. Research on political information and social media: Key points and challenges for the future. 964–974. Accessed: 26-12-2023.

[2] E. Kestilä-Kekkonen and Å. von Schoultz. 2020. Ehdokkaat vaalikentillä: Eduskuntavaalit 2019. In 2490–0680. 2020, Oikeusministeriö. Access date: 10.5.2022.

[3] H.-C. Lin. 2017. How political candidates' use of Facebook relates to the election outcomes. *Int. J. Market Res.* 59, 1 (2017), 77–96.

[4] C. B. Williams and G. J. J. Gulati. 2013. Social networks in political campaigns: Facebook and the congressional elections of 2006 and 2008. *New Media Societ.* 15, 1 (2013), 52–71.

[5] A. Jungherr. 2016. Twitter use in election campaigns: A systematic literature review. *J. Inf. Technol. Polit.* 13, 1 (2016), 72–91.

[6] M. Magin, N. Podschuweit, J. Hassler, and U. Russmann. 2017. Campaigning in the fourth age of political communication. A multi-method study on the use of Facebook by German and Austrian parties in the 2013 national election campaigns. *Inf., Commun. Societ.* 20, 11 (2017), 1698–1719.

[7] K. Strandberg. 2013. A social media revolution or just a case of history repeating itself? The use of social media in the 2011 Finnish parliamentary elections. *New Media Societ.* 15, 8 (2013), 1329–1347.

[8] X. Zhang. 2018. Social media popularity and election results: A study of the 2016 Taiwanese general election. *PloS One* 13, 11 (2018), e0208190.

[9] J. Murphy, M. W. Link, J. H. Childs, C. L. Tesfaye, E. Dean, M. Stern, J. Pasek, J. Cohen, M. Callegaro, and P. Callegaro. 2014. Social media in public opinion research: Executive summary of the AAPOR task force on emerging technologies in public opinion research. *Pub. Opin. Quart.* 78, 4 (2014), 788–794.

[10] M. F Schober, J. Pasek, L. Guggenheim, C. Lampe, and F. G. Conrad. 2016. Social media analyses for social measurement. *Pub. Opin. Quart.* 80, 1 (2016), 180–211.

[11] F. Diaz, M. Gamon, J. M. Hofman, E. Kıcıman, and D. Rothschild. 2016. Online and social media data as an imperfect continuous panel survey. *PloS One* 11, 1 (2016), e0145406.

[12] R. M. Groves. 2011. Three eras of survey research. *Pub. Opin. Quart.* 75, 5 (2011), 861–871.

[13] A. Ceron, L. Curini, S. M. Iacus, and G. Porro. 2014. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media Societ.* 16, 2 (2014), 340–358.

[14] V. Pekar, H. Najafi, J. M. Binner, R. Swanson, C. Rickard, and J. Fry. 2022. Voting intentions on social media and political opinion polls. *Govern. Inf. Quart.* 39, 4 (2022), 101658.

[15] D. Rousidis, P. Koukaras, and C. Tjortjis. 2020. Social media prediction: A literature review. *Multim. Tools Applic.* 79, 9 (2020), 6279–6311.

[16] D. Gayo-Avello. 2013. A meta-analysis of state-of-the-art electoral prediction from Twitter data. *Soc. Sci. Comput. Rev.* 31, 6 (2013), 649–679.

[17] K. Brito, R. L. C. Silva Filho, and P. J. L. Adeodato. 2021. A systematic review of predicting elections based on social media data: Research challenges and future directions. *IEEE Trans. Computat. Soc. Syst.* 8, 4 (2021), 819–843.

[18] P. Chauhan, N. Sharma, and G. Sikka. 2021. The emergence of social media data and sentiment analysis in election prediction. *J. Amb. Intell. Human. Comput.* 12, 2 (2021), 2601–2627.

[19] L. Phillips, C. Dowling, K. Shaffer, N. O. Hodas, and S. Volkova. 2017. Using social media to predict the future: A systematic literature review. arXiv preprint arXiv:1706.06134.

[20] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. 2011. Election forecasts with Twitter: How 140 characters reflect the political landscape. *Soc. Sci. Comput. Rev.* 29, 4 (2011), 402–418.

[21] T. Vepsäläinen, H. Li, and R. Suomi. 2017. Facebook likes and public opinion: Predicting the 2015 Finnish parliamentary elections. *Govern. Inf. Quart.* 34, 3 (2017), 524–532.

[22] A. Jungherr, P. Jürgens, and H. Schoen. 2012. Why the pirate party won the German election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, to, sander, pg, & welpe, im "predicting elections with twitter: What 140 characters reveal about political sentiment". *Soc. Sci. Comput. Rev.* 30, 2 (2012), 229–234. https://journals.sagepub.com/doi/10.1177/0894439311404119

[23] D. Gayo-Avello. 2012. "I wanted to predict elections with Twitter and all I got was this lousy paper"—A balanced survey on election prediction using Twitter data. *arXiv preprint* arXiv:1204.6441.

[24] P. Rita, N. António, and A. P. Afonso. 2023. Social media discourse and voting decisions influence: Sentiment analysis in tweets during an electoral period. *Soc. Netw. Anal. Min.* 13, 1 (2023), 46.

[25] K. Brito and P. J. L. Adeodato. 2022. Measuring the performances of politicians on social media and the correlation with major Latin American election results. *Govern. Inf. Quart.* 39, 4 (2022), 101745.

[26] G.-J. Put, J. Smulders, and B. Maddens. 2019. How local personal vote-earning attributes affect the aggregate party vote share: Evidence from the Belgian flexible-list PR system (2003–2014). *Politics* 39, 4 (2019), 464–479.

[27] D. J. Ruck, N. M. Rice, J. Borycz, and R. A. Bentley. 2019. Internet research agency Twitter activity predicted 2016 US election polls. *First Mond.* 24, 7 (2019).
[28] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer. 2019. Fake news on Twitter during the 2016 US presidential election. *Science* 363, 6425 (2019), 374–378.
[29] F. Giglietto, N. Righetti, and G. Marino. 2019. Diverging patterns of interaction around news on social media: Insularity and partisanship during the 2018 Italian election campaign. *Inf., Commun. Societ.* 22, 11 (2019), 1610–1629.
[30] D. Rofrío, A. Ruiz, E. Sosebee, Q. Raza, A. Bashir, J. Crandall, and R. Sandoval. 2019. Presidential elections in Ecuador: Bot presence in Iwitter. In *6th International Conference on eDemocracy & eGovernment (ICEDEG'19)*. IEEE.
[31] V. do Nascimento Silva and R. H. A. Silva. 2019. Are algorithms affecting the democracy in Brazil. In *International Symposium on Ethical Algorithms*.
[32] M. A. M. Nizah and A. R. A. Bakar. 2019. Whatsapp election in malaysia: Assessing the impact of instant messaging on malaysia's 14th general election. *International Journal of Academic Research in Business and Social Sciences* 9, 3 (2019), 132–146.
[33] J. S. Santos, F. Bernardini, and A. Paes. 2021. A survey on the use of data and opinion mining in social media to political electoral outcomes prediction. *Soc. Netw. Anal. Min.* 11 (2021), 1–39.
[34] M. C. MacWilliams. 2015. Forecasting congressional elections using Facebook data. *PS: Polit. Sci. Polit.* 48, 4 (2015), 579–583.
[35] K.-C. Chang, C.-F. Chiang, and M.-J. Lin. 2021. Using Facebook data to predict the 2016 US presidential election. *Plos One* 16, 12 (2021), e0253560.
[36] J. B. Kristensen, T. Albrechtsen, E. Dahl-Nielsen, M. Jensen, M. Skovrind, and T. Bornakke. 2017. Parsimonious data: How a single Facebook like predicts voting behavior in multiparty systems. *PloS One* 12, 9 (2017), e0184562.
[37] R. Sandoval-Almazan and D. Valle-Cruz. 2020. Sentiment analysis of Facebook users reacting to political campaign posts. *Digit. Govern.: Res. Pract.* 1, 2 (2020), 1–13.
[38] H. T. Le, G. Boynton, Y. Mejova, Z. Shafiq, and P. Srinivasan. 2017. Revisiting the American voter on Twitter. In *CHI Conference on Human Factors in Computing Systems*.
[39] L. Hagen et al. 2022. Rise of the machines? Examining the influence of social bots on a political discussion network. *Soc. Sci. Comput. Rev.* 40, 2 (2022), 264–287.
[40] R. R. Lau and D. P. Redlawsk. 2006. *How Voters Decide: Information Processing in Election Campaigns*. 2006: Cambridge University Press. 255–264.
[41] M. M. Skoric, J. Liu, and K. Jaidka. 2020. Electoral and public opinion forecasts with social media data: A meta-analysis. *Information* 11, 4 (2020), 187.
[42] V. Isotalo, M. Mattila, and Å. von Schoultz. 2020. Ideological mavericks or party herd? The effect of candidates' ideological positions on intra-party success. *Elector. Stud.* 67 (2020), 102187.
[43] Å. Von Schoultz and A. Papageorgiou. 2021. Policy or person? The electoral value of policy positions and personal attributes in the Finnish open-list system. *Party Polit.* 27, 4 (2021), 767–778.
[44] H. S. Christensen, M. S. L. Rosa, and K. Grönlund. 2020. How candidate characteristics affect favorability in European parliament elections: Evidence from a conjoint experiment in Finland. *Eur. Union Polit.* 21, 3 (2020), 519–540.
[45] M. L. McDermott. 1998. Race and gender cues in low-information elections. *Polit. Res. Quart.* 51, 4 (1998), 895–918.
[46] H. Coffé and E. Theiss-Morse. 2016. The effect of political candidates' occupational background on voters' perceptions of and support for candidates. *Polit. Sci.* 68, 1 (2016), 55–77.
[47] B. Maddens, B. Wauters, J. Noppe, and S. Fiers. 2006. Effects of campaign spending in an open list PR system: The 2003 legislative elections in Flanders/Belgium. *West Eur. Polit.* 29, 1 (2006), 161–168.
[48] T.-h. Tsai and J. Gill. 2013. Interactions in generalized linear models: Theoretical issues and an application to personal vote-earning attributes. *Soc. Sci.* 2, 2 (2013), 91–113.
[49] P. Suhonen. 2006. *Mielipidetutkimukset ja yhteiskunta*. Tampere University Press.
[50] O. S. O. Finland. 2019. Official Statistics of Finland: Parliamentary elections [e-publication). Retrieved from http://www.stat.fi/til/evaa/index_en.html. Retrieved 15 April 2019, from Official Statistics of Finland http://www.stat.fi/til/evaa/index_en.html
[51] Yle. Welcome to Yle's election compass! Retrieved from: https://vaalikone.yle.fi/eduskuntavaali2019/?lang=en-EN
[52] T. Hastie, R. Tibshirani, and J. H. Friedman. 2009. The elements of statistical learning: Data mining. *Infer. Predict.* 2 (2009). Springer.
[53] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, Ron Weiss, and V. Dubourg. 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12 (2011), 2825–2830.
[54] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li. 2005. *Applied Linear Statistical Models*. McGraw-Hill.
[55] S. Seabold and J. Perktold. 2010. StatsModels: Econometric and statistical modeling with Python. In *9th Python in Science Conference*.
[56] U. Grömping. 2015. Variable importance in regression models. *Wiley Interdisc. Rev.: Computat. Stat.* 7, 2 (2015), 137–152.
[57] R. M. O'Brien. 2007. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* 41 (2007), 673–690.

Received 9 October 2023; revised 12 February 2024; accepted 25 February 2024