Can we predict multi-party elections with Google Trends data? Evidence across elections, data windows, and model classes

Jan Behnert[1], Dean Lajic[1] and Paul C. Bauer[2,3]*

*Correspondence:
mail@paulcbauer.de

[1] MZES, University of Mannheim, Mannheim, Germany
[2] Department of Political Science, University of Freiburg, Freiburg, Germany
[3] Department of Statistics, LMU Munich, Munich, Germany

Abstract

Google trends (GT), a service aggregating search queries on Google, has been used to predict various outcomes such as as the spread of influenza, automobile sales, unemployment claims, and travel destination planning [1, 2]. Social scientists also used GT to predict elections and referendums across different countries and time periods, sometimes with more, sometimes with less success. We provide unique evidence on the predictive power of GT in the German multi-party systems, forecasting four elections (2009, 2013, 2017, 2021). Thereby, we make several contributions: First, we present one of the first attempts to predict a multi-party election using GT and highlight the specific challenges that originate from this setting. In doing so, we also provide a comprehensive and systematic overview of prior research. Second, we develop a framework that allows for fine-grained variation of the GT data window both in terms of its width and distance to the election. Subsequently, we test the predictive accuracy of several thousand models resulting from those fine-grained specifications. Third, we compare the predictive power of different model classes that are purely GT data based but also incorporate polling data as well as previous elections. Finally, we provide a systematic overview of the challenges one faces in using GT data for predictions part of which have been neglected in prior research.

Introduction

Google trends (GT), a service aggregating search queries on Google, has been used to predict various outcomes such as the spread of influenza, automobile sales, unemployment claims, and travel destination planning [2]. These sparked the interest of political scientists who subsequently used GT data to predict elections with binary outcomes, e.g., presidential elections in the US or referenda such as the Brexit referendum, and, often claim to be successful [3]. In principle, Google Trends could provide a cheap data

Behnert *et al. Journal of Big Data*   (2024) 11:30

Page 2 of 21

source that may extend previous predictive models that rely on polling data as well as structural data such as previous election results [4]. Exploring new data sources such as GT is also warranted given the persisting discussion on the validity of polling data [5].[1]

In our study we pursue the following research question: Can we use Google Trends data to predict election results in a multi-party system? Thereby, we make a series of contributions to scholarship using GT for predictive purposes generally but also specifically for elections: First, we are among the first to provide evidence on the predictive power of GT in the multi-party system setting. Thereby, we use GT to predict four elections in Germany and highlight the specific challenges that originate from the multi-party context. We also provide a systematic review of previous research highlighting variation across several important dimensions. This review both helps us to highlight our contributions and can represent a starting point for future research on GT predictions (see Table 1).

Second, when using GT for predictions one of the most important choices lies in the GT data window on which those predictions are based. The data window may vary in terms of width, e.g., our prediction could be based on averaging 1 week of GT data as opposed to 3 weeks of GT data. And, the data window may vary in terms of distance to the event that shall be predicted, e.g., we could predict an election using a data window that ends just two days before the election or a data window that ends three months before the election. Granka [6] suggested to explore moving averages to assess how well models withstand changes closer to the election date. While previous studies have varied these aspects, we are the first to build a framework that allows to cycle through fine-grained values of both width and distance testing the predictive accuracy of thousands of resulting models.

Third, following previous studies, we compare the predictive power of different model classes. Besides, a model that only includes GT data, we explore the predictive accuracy of models that combine GT data with election data and polling data. And, we compare our predictions to simple polling data. This provides us with an answer as to whether GT really does represent an alternative to classic purely poll-based methods and whether combinations are fruitful.

Fourth, we provide a systematic overview of the challenges one faces in using GT data for predictions part of which have been neglected in prior research. These comprise choices on the GT platform, e.g., we can restrict that data to searches belonging to certain categories, but also the varying nature of GT data across samples. We discuss which previous studies have acknowledged these and other challenges and tackle them in a systematic fashion to study their impact. Thereby, we provide a blueprint for future GT research.

In Section "Using Google Trends to predict elections and other phenomena" we start by providing an overview of research that leverages GT data for predictive purposes generally but also for political outcomes. In Section "Data and Methods" we explain our methodological approach, the data we are collecting and using as well as the predictive models we are building. Section "Results" presents our results namely the predictions

---

[1] More recently, scholars have turned to using reweighted non-representative polls to predict elections (e.g., [5]).

Behnert *et al. Journal of Big Data*   (2024) 11:30

Page 3 of 21

Table 1 Overview of choices in previous studies

Study	Election(s)	Width(s)	Distance	Categories used	Data type used	Multiple GT datasets	Explain / mention search term selection
Peer-reviewed							
Raubenheimer et al. [22]	2020 New Zealand cannabis referendum	1 week; 3 months	1 day before event	No	Hourly daily	Yes	Yes
Prado-Román et al. [3]	2004-2016 presidential elections US and 2004–2019 presidential elections Canada	1/2/3 month(s)	1 day before event	?	Daily	No	No
Mavragani and Tsagarakis [18]	The 2014 Scottish the 2015 Greek the 2016 UK Hungarian Italian and the 2017 Turkish Referendum	1 week; 1 month	1 day before event	Yes (no further explanation why)	Hourly daily	No	No
Mavragani and Tsagarakis [23]	Greek referendum 2015	8 days; 1 day; 12 hours	1 day before event & N= 0; on Referendum day	?	Hourly	No	Yes
Polykalas et al. [19]	2005-2013 German Elections (only parties CDU/SPD)	1 month	1 day before event	?	Daily	No	Yes
Harkan and Eryanto [24]	2019 Indonesian Presidential Election	23 Sept 2018 - 16 May 2019	1 day before event	No	weekly?	No	No
Polykalas et al. [14]	Greek elections 2007-2012 & Spanish national elections 2008-2011	1 month	1 day; 1/2 week(s)	?	Daily	No	Yes
Yasseri and Bright [25]	Iranian election 2013 German election 2013 UK election 2010	?	?	?	Daily?	No	No
Granka [6]	US Presidential Elections 4' 8' and 12'	1 month	2 months	No	weekly	No	No
Vergara-Perucich [26]	Chile: Presidential election 2006–2021	121 days	5 days	?	Daily	No	No
Non-peer reviewed							
Sjövill [20ª]	Swedish general election 2010-2018	1/2 weeks; 1 month	1 day before event	No	Daily	No	Yes
Wolf [15]	US Presidential Elections 16' swing states	2/4 Days	Depending on poll data	No	Daily hourly?	No	?
Askitas [17]	Greek referendum 2015	1 week	1 day before event	No	Hourly	No	Yes
Askitas [16]	Irish "Gay Marriage" Referendum 2015	1 week	1 day before event	No	Hourly	No	Yes

ª = Non-binary outcome; ? = information not available/provided

of four elections. In the conclusion we summarize the most important insights (see Section "Conclusion").

Using Google Trends to predict elections and other phenomena

The proportion of internet users aged 14 and over in Germany has risen from 37% in 1991 to 91% individuals aged 14 and older [7]. In terms of search engines Google is by far the dominant search engine with market shares of 80.4% (desktop) and 96.8% (mobile) [8]. Nonetheless, as we will discuss below, Google users are not necessarily representative of the German electorate.

Predicting phenomena with Google Trends

Google has made its data on Google searches freely available for everyone in the year 2006 (see https://trends.google.com/). Rather than absolute numbers of searches, GT provides data on interest in a search term relative to all other search terms in a country or region over a selected period of time. GT data comes along with certain advantages such as cost-free access to aggregated big data, a sample that is (ideally) representative of all Google searches or an unfiltered real-time sample. Soon after going public GT made headlines and the number of studies using GT data has grown significantly. In general, the assumption underlying those studies is that search queries in Google reflect the genuine interests or intentions of people. While GT was most frequently used in the field of Computer Science, usage in other disciplines has picked up [9]. In a ground-breaking study Ginsberg and colleagues used GT data and its real-time nature to predict the spread of influenza, comparing its accuracy to predictions by the Government Agency Center for Disease Control and Prevention (CDC) [2]. This first study started a debate on the advantages but also deficiencies of GT flu predictions [10–12]. The enormous potential of GT data was also demonstrated by Choi & Varian [1], who predicted economic indicators including automobile sales, unemployment claims, travel destination planning, etc. The continued popularity is reflected in recently published papers that used GT data in the context of the COVID-19 pandemic. Brodeur et al. [13] for example examined whether COVID-19 and the associated lockdowns initiated in Europe and America led to changes in well-being.

Predicting elections with Google Trends

Table 1 summarizes studies that have used GT data to predict elections. Reviewing the literature a few aspects stand out. First, the focus usually lies on binary electoral outcomes. Prado-Román et al. [3], for example, were able to predict the final results of all presidential elections in the US and Canada for the time period 2004 – 2016. Polykalas et al. [14] were able to forecast Greek and Spain election results. However, other studies found GT to be less helpful in predicting elections, despite their focus on settings with binary outcomes [15]. Other studies illustrated that GT data can be used to predict referendum results [16, 17]. Mavragani & Tsagarakis [18] successfully predicted six referendums in different countries in Europe between 2014 and 2016, for example the Brexit referendum. Polykalas et al. [19] used GT to predict elections in Germany, examining only the two most popular parties, the SPD and the CDU, trying to predict which of the two parties will win. In contrast to the previously mentioned

studies, the authors additionally weight their GT predictions using previous election results to control for the selection bias, i.e., the fact that not everyone uses the internet. As indicated in Table 1, one of the few exceptions predicting multi-party elections is [20]. Sjövill's thesis uses GT data to predict party shares focusing on the three Swedish federal elections 2010, 2014 and 2018. Sjövill [20] emphasizes the importance of weighting the GT predictions using previous final election results and polling data, to control for the sample selection bias of GT data. Sjövill [20] compares different models with the different weighting methods and finds that they mostly have the same predictive accuracy as the average of pre-election opinion polls. The weighting method using actual polling data proved to be the most informative. Second, the GT data window that is used for the prediction is defined through its width, i.e., the end time minus the start time of the window, and the distance to the event that shall be predicted. Across studies there is strong variation in terms of the GT data windows chosen both in terms of width and distance (cf. Table 1). Most studies pick one to three different widths, usually one week or one month. And, in terms of distance studies usually pick a data window that ends just before the election. Naturally, the question arises whether the findings made across these studies would be the same if we were to vary the width and distance of the underlying GT data. It is one of our aims to provide a more systematic approach to choosing the data window comparing predictions for a wide variety of choices. Third, while researchers have compared GT predictions to classical polling data, they have also explored different weighting schemes that may decrease sample selection bias [20]. Sample selection bias is present if, e.g., voters of conservative parties are on average older and thus use the internet less. As a consequence they are underrepresented among Google users leading to an underestimation of the vote share of conservative parties. Polykalas et al. [19] weighted their GT predictions with election results from the previous election. Sjövill [20] constructed three different models: A long-term model that weighs GT predictions with partys' previous election results; an intermediate model weighting the GT predictions using semi-annual and highly representative polling data from a respected election poll in Sweden; and a short-term model which used average monthly polling data for weighting. While all models came close to the results of the election polls in the election years studied, the short time model proved to be the most informative. Inspired by this research we also compare different weighting schemes as described in Section "Prediction models and benchmark" Finally, with few exceptions, most studies remain silent on a set of important characteristics of the GT platform and its data. At the same time these characteristics can affect any predictive exercise. A first issue is that data provided by GT represents a random sample that changes over time [21]. As summarized in Table 1, column "Multiple GT datasets", we only found one study that bases its predictions on several GT data samples [21]. Following Raubenheimer et al. [21] we develop a systematic sampling strategy for GT data and average our predictions across those samples as described in Section "Collecting GT data" and Additional file 1: Section S2. A second issue is the selection of GT search terms (see Section "Search terms and category filter"). This selection potentially has the strongest effect on any predictions we make, hence, transparently communicating how and following which rational these terms have been selected is of utmost importance. A third issue is whether to further refine

Behnert *et al. Journal of Big Data*   (2024) 11:30

Page 6 of 21

search terms by picking a category filter provided by Google (see Section "Category filter"). Such filters attempt to identify searches that belong to particular topics helping to identify only relevant searches. With the exception of Mavragani & Tsagarakis [22] no previous studies made use of such categories filters (cf. column "Cat. used" in Table 1). Below, we compare the impact of basing predictions on searches refined through such a category and non-refined searches.

Data and methods

In our analysis we rely on three types of data, Google trends data (see Section "Collecting GT data") as well as data on polls and actual election results (see Section "Collecting polling and election data").

Google Trends as a data source

GT provides access to a largely unfiltered sample of "real-time" searches on different topics (up to 36 h before the actual time you conduct the search) or a filtered and representative sample (as claimed by Google) of searches that are older than 36 h starting from the year 2004. The data can be obtained for different search types that correspond to different Google products like "Web search," "News," "Images," "Shopping" and "Youtube". Importantly, GT does not provide access to data on individual searches. Rather the data is anonymized and Google aggregates the data to the federal state level, country level or world level. Besides, it is possible to filter searches belonging to different categories, e.g., "Law and Government", with the aim of only getting searches for the word's meaning one is interested in. The result we get are a standardized, relative measure of search volume for a single word/search term, a combination of search terms using operators,[2] or comparisons, i.e., one input in relation to the other inputs, [3]over a selected time period. Google calculates how much search volume in each region a search term or query had, relative to all searches in that region. Using this information, Google assigns a measure of popularity to search terms (scale of 0–100), leaving out repeated searches from the same person over a short period of time and searches with apostrophes and other special characters [27]. The maximum of the scale corresponds to a search term's maximum level of popularity relative to other search terms and time periods.

Search terms and category filter

The first step in using Google searches to predict phenomena is the selection of search terms on which we base our predictions. Since words may have multiple meanings, e.g., jaguar could be an animal or a car, GT provides a category filter to get data for the right version of the word. Only one previous study made use of the category filter in Google Trends [18], a shortcoming since categories help purge the search term for multiple meanings and thus assure that one gets results for the right version of that search term. For instance, in our case adding the category "Law and Government" when searching for

---

[2] Available Operators: No quotation marks (results for each word in your query), Quotation marks (coherent search phrase), Plus sign (serves as function of an OR-operator) and Minus sign (Excludes word after the operator) (**Google Trends Help 2023d**).

[3] Possible to compare up to five groups of terms at one time and up to 25 terms in each group (Google Trends Help 2023a).

the search term "Grüne" assures that the political party is meant and not the color [28]. Below we describe how we selected the search terms and which selections we chose for the GT's category filter.

Category filter

The first step in our analysis was to compare all supercategories within Google's "Web Search" product across our four elections (09', 13', 17' and 21'), to select the ideal category for our analysis.[4] For this purpose, we searched for the abbreviations of the major political parties (e.g., CDU, SPD, etc.) using a time window of January 1st to September 26th for all four elections and setting geographic location to Germany. Thereby, it became apparent that the Law & Government category was the most suitable and also the most reasonable for our purposes, as it places search queries in the political context. Previous studies have mostly chosen the "All categories" category which performs significantly worse than "Law & Government" in the German context. We refer the reader to Additional file 1: Section S1 for the corresponding analyses.

Search terms

Fundamentally, we assume that google searches for political parties and politicians may reflect vote intentions and choice. Therefore, we can use these searches to predict the latter. Importantly, in pursuing the steps described below we always filter searches using the category described above. Table 2 depicts the final search queries that we identified following the steps below. In Step 1, we try out different search terms and explore their popularity over the entire period of four years until the election as depicted in Fig. 1. If a search term reaches a peak shortly before the election, we assume that this search interest peaking before an election is an indicator which represents an election intention. In Fig. 1 an example is the search term "CDU" and "Angela Merkel" as well as "Armin Laschet" in the 21' election after Angela Merkel resigned. Following this process, we found that across the different parties, the query of the party abbreviation, the full party name and the respective top candidate had peak values before the election. Figure S2 provides more examples for the other parties.

In Step 2, we compared all search terms for a party identified in Step 1 to explore how relevant the single search terms were in relation to the other search terms. The comparison showed that the party abbreviation, e.g., SPD, is extremely well suited whereas the full party name is not. Special cases were the parties "Die Grünen" and "Die Linken". In the case of the party "Die Grünen", the search query "Grüne" delivered the best results, since the use of a category assures that only search terms in the political context are used and the term already includes longer and other spellings such as "Die Grünen" or "Bündnis 90 die Grünen". As depicted in Fig. 1, the same can be observed for the party "Die Linke", for which the search query "Linke" is best suited. In addition, using the same strategy as above we found that the names of the respective leading party candidates yielded even better predictions for all parties. In

---

[4] Supercategories: All categories, Arts & Entertainment, Autos & Vehicles, Beauty & Fitness, Books & Literature, Business & Industrial, Computer & Electronics, Finance, Food & Drink, Games, Health, Hobbies & Leisure, Home & Garden, Internet & Telecom, Jobs & Education, Law & Government, News, Online Communities People & Society, Pets & Animals, Real Estate, Reference, Science, Shopping, Sports, Travel.
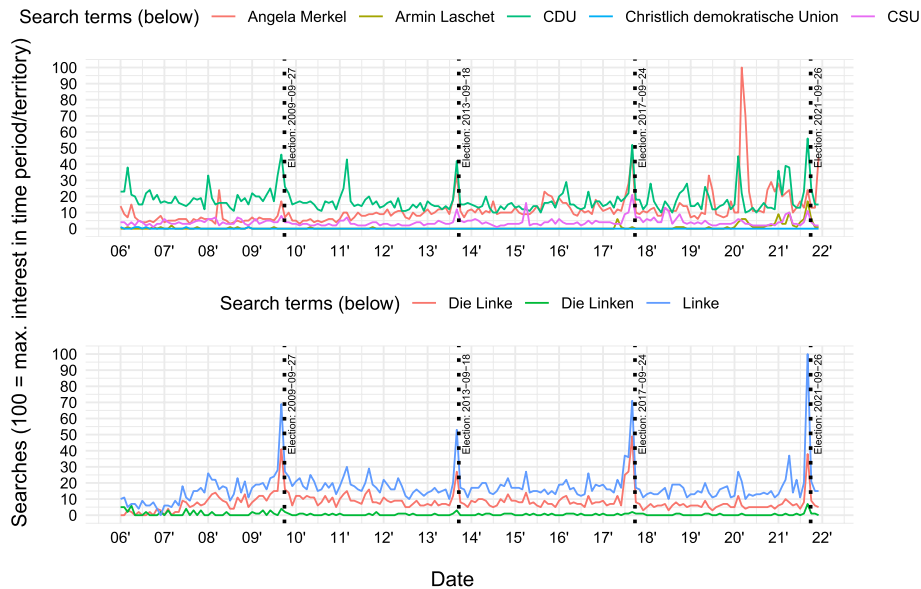
Fig. 1 Search queries for the parties (CDU and Die Linke)

Table 2 Final search queries

Election date	Search queries
2009-09-27	CDU + CSU + Angela Merkel; SPD + Frank Walter Steinmeier; Grüne + Jürgen Trittin + Renate Künast; Linke + Gregor Gysi; FDP + Guido Westerwelle
2013-09-22	c("CDU + CSU + Angela Merkel", "SPD + Peer Steinbrück", "Jürgen Trittin + Katrin Göring Eckardt + Grüne + ", "Linke + Sarah Wagenknecht + Gregor Gysi", "FDP + Philipp Rösler"); c("Bernd Lucke + Afd", "CDU + CSU + Angela Merkel", "SPD + Peer Steinbrück", "Jürgen Trittin + Katrin Göring Eckardt + Grüne", "FDP + Philipp Rösler")
2017-09-24	c("CDU + CSU + Angela Merkel", "SPD + Martin Schulz", "Cem Özdemir + Katrin Göring Eckardt + Grüne", "Linke + Sarah Wagenknecht + Dietmar Bartsch", "FDP + Christian Lindner"); c("Alice Weidel + Alexander Gauland + Afd", "CDU + CSU + Angela Merkel", "SPD + Martin Schulz", "Cem Özdemir + Katrin Göring Eckardt + Grüne", "FDP + Christian Lindner")
2021-09-26	c("CDU + CSU + Armin Laschet", "SPD + Olaf Scholz", "Annalena Baerbock + Grüne", "Linke + Janine Wissler + Dietmar Bartsch", "FDP + Christian Lindner"); c("Alice Weidel + Tino Chrupalla + Afd", "CDU + CSU + Armin Laschet", "SPD + Olaf Scholz", "Annalena Baerbock + Grüne", "FDP + Christian Lindner")

the case of dual leadership, both candidates were added.[5] The final search queries of a party are made up of the party abbreviation and the top candidate(s), which are linked with the "+" operator. This functions according to the scheme of an OR operation, i.e., it sums up the search interest for each individual term. In step 3, we compared the predictive power of a set of search terms containing both the party abbreviation and the top candidate with another set containing only the party abbreviations. We

---

[5] A special case was the party "Die Linke," which nominated eight top candidates in 2013. By comparing the search queries of the individual top candidates separately, we filtered out the only two top candidates that seemed relevant for the prediction, who were "Gregor Gysi" and "Sarah Wagenknecht." For the AfD we used the search term "Afd" finding that it to be equivalent to "AFD" and "AfD".

Behnert *et al. Journal of Big Data*   (2024) 11:30

Page 9 of 21

found that the additional consideration of the top candidates provided significantly better results.[6] Table 2 displays the final search queries.

Collecting GT data

To collect data on GT, we use the R package "gtrendsR" (Massicotte & Eddelbuettel, 2023). For the purpose of our analysis we rely on the comparison function of GT. In the following, the term "request(s)" implies that we collect data from GT using this function. Unfortunately, Google Trends only allows us to compare up to five groups of terms at a time. In Germany, however, the total number of major political parties in Germany amounts to six since the appearance of the party AfD in 2013. For the respective elections (2013–2021) we conducted two requests: The first request included our search queries for the major political parties CDU, SPD, FDP, "Die Grünen" and "Die Linke." The second request is similar to the first one, with the difference that we exchanged "Die Linke" with AfD, in order to get data for the AfD. Since, the relative scale for the comparative GT data is always anchored (setting the maximum) using the most popular search term, we can leave out or exchange search terms as it does not change their popularity relative to the maximum. As a result we get estimates of search interest for all six parties. Table 2 displays the final search queries. We collected Google Trends data, having set the geo-location to Germany, for our search queries from the first day of an election year until the 26th September for 2009, 2013, 2017 and 2021. Additionally we collected data from the election year 2005, which we later used to construct a weighting factor.[7]

Google Trends samples

Importantly, Google Trends draws new samples of all searches on the platform several times a week [21]. As a result, the data varies slightly from sample to sample. Google states that the samples taken are representative of all Google search queries. However, Google does not provide any information on how this representativeness is achieved [27]. Since Google does not specify at what intervals new samples are taken, we collected new data sets every hour for several weeks and compared them. It turned out that new samples were taken at least once a day, sometimes more often. To account for the variation one may find across samples, we collect GT data at 10 different time points between 01/12/22 and 10/12/22 (see Additional file 1: Table S1 in Section S2). We compare datasets across time points to assure that we base our estimates on non-identical datasets as

---

[6] Regarding the number of terms within a search query for a party, one could argue that the result of a search query should be divided by the number of search terms, because the search queries of some parties contain more individual terms and could thus be overrepresented (e.g. if a party has a dual leadership). In our opinion, this argumentation is not plausible, because the search terms for one party never provide an equal share of the search interest. The share of the term for the party abbreviation is much higher than that for the party candidates, so it is illogical to treat them equally. We checked this scenario by searching only for the more popular candidate when there is a dual leadership, which results in the same amount of search terms for each party. This, however, leads to much worse predictions than our main analysis. Thus, we assume that the parties are not overrepresented by their multiple top candidates.

[7] Choosing a period longer than 270 days for the election our years, results in getting weekly data instead of daily data. One might also ask if it would make a difference to pull one dataset that covers all interval and distance combinations, as in our case, or to pull single datasets for each interval and distance combination. We checked this and it made no difference since we only use the comparison function of Google Trends. Accordingly, the ratio of the parties to each other remains the same [29].
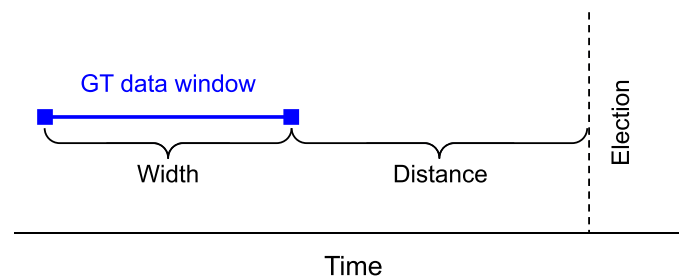
Fig. 2 Logic of Google Trends data windows

to account for sampling error. In our data analysis, we then use the average values across those datasets as well as the associated confidence intervals. Surprisingly, we are the first study within the election prediction literature to acknowledge this problem, although it is an obvious source of error [29]. Additional file 1: Figure S2 in Section S2 visualizes the variation of the 10 GT datasets. Moreover, in contrast to statements in the literature referring to data from the Google Trends website [30] and the GTEH API [22], we find that sometimes more than one different sample can be obtained on the same day (see Additional file 1: Table S1 in Section S2) and that drawing samples from different PCs in different networks at the same time yields the same sample when using gtrendsR.

Google trends data windows

After having collected data from 1 January to 27 September of each election year, the question arises which data windows within these data are best suited for election prediction. As depicted in Fig. 2, data windows are defined by their width (length of time period) and distance to the election. In previous research authors rarely provided a justification for the corresponding choices.

In what concerns width, the majority of studies used time window(s) with a width of 1/2 weeks or 1/3 months. In contrast, our aim is to systematically examine as many data windows as possible in order to find the best possible prediction windows and to provide guidance for future research. In doing so we compare small with large data windows. Small data windows better reflect the searches occurring in short time spans and the current mood in those time periods. Wider data windows reflect average interest across a longer time period and average out non-meaningful spikes (e.g., TV appearances). We compare data windows of eight widths namely 7, 14, 21, 28, 42, 56, 70 and 91 days.

In what concerns distance, previous studies have barely examined this aspect and have mostly chosen a distance of one day to the election. However, election campaign leaders and demoscopists will be more interested in how parties are predicted to perform at an earlier timepoint, e.g., two weeks or three months before the election. We compare the predictive power of GT data windows of distance one to 150 days before election day. For instance, a distance of, e.g., 30 days means that the GT data window ends 30 days before the election. In total, we compare 8 (number of widths) * 150 (number of distances) = 1200 different GT data windows per election.

Collecting polling and election data

In addition to Google Trends data we also collected polling data both as a comparison for our Google Trend predictions but also to weight our Google Trends predictions. The polling data comes from one of the most reputable German polling institutes, Infratest dimap. Finally, the outcome we want to predict are actual election results. We collected the corresponding data from the official source of the federal government (Bundeswahlleiter 2023).

Prediction models and benchmark

In our analysis, we predict party shares in four federal elections in Germany: 2009, 2013, 2017 and 2021. We build three broad classes of predictive models:

Our first class of models, MC1, consists of the raw, unweighted Google Trends data; The second class of models, MC2, uses a weighting factor based on the results of the preceding election; Our last class of models, MC3, weights the GT predictions using polling results of Infratest dimap, which will be explained in more detail below. In total we estimate 4 (elections) * 3 (model classes) * 1200 (data windows) = 14,400 models. We compare predictions of models based on GT data to predictions based on models drawing on polls for the time windows to benchmark their performance.[8] For each election, we compare our predictions for the single parties with the actual election results. Below we visualize and evaluate both errors across our models for single parties $y_p - \hat{y}_p$ as well as averaged across parties using the $MAE = \frac{\sum_{i=1}^{n} |y_p - \hat{y}_p|}{n}$ where p corresponds to a party and y_p and $\hat{y}_p$ to the true and predicted party share respectively.

MC1 models predict party shares solely with Google Trends data. To obtain these we proceeded as follows: We started by calculating the Google proportion for each party p for a data window i. As shown in Eq. 1, in the first step, we use the average Google search interest of each party p for the examined data window i, for example, the average search interest for the search query of the CDU for the data window with width 7 days and distance 7 days. Then we divide it by the sum N of all parties average search interest for that data window. The Google proportion thus serves as a prediction in percent for the respective party.

$$\text{Google Proportion}_{p,i} = \frac{\text{Avg. Google search interest}_{p,i}}{\sum_{p=1}^{N} \text{Avg. Google search interest}_{p,i}} \tag{1}$$

If we add up the Google proportions of all the parties examined, we arrive at 100%. This Google proportion serves as the sole basis for models of class MC1. Highlighting our strategy reveals a slight disadvantage of our analysis data compared to the election polls. While the election polls also include the category "Other," which in extreme cases can go up to 8.7% of the votes, we lack this category on GT. This means that in the polls the main parties can sometimes only get 91.3%, whereas in our analysis they get 100%. As a result, we assume that party shares are slightly overestimated in the GT data. We provide additonal analyses constructing an "other" category from google search data in

---

[8] The term model does not necessarily refer to a sophisticated statistical model here. For instance, the models based on polls are simply the party shares as measured in the polls.

Additional file 1: Section S7 finding that the results do not change much. In our conclusion we discuss possible ways of how the "other" category could be accounted for when collecting GT data.

For models of class MC2 and MC3 we use weighting factors in conjunction with the party shares predicted by GT proportions. Models of class MC2 are based on the approach of Polykalas et al. [14], who used the results of the previous federal election to calculate a weighting factor. As shown in *Eq. 2*, to calculate the weighting factor *WMC*2 for a party *p*, for the data window *i* and the election year *T*, we divide the previous election results of that party *p* of the previous election year $T - 4$ by the respective Google proportion. For example, to predict the SPD's share in 2017 we must first calculate the weighting factor *WMC*2. For that, we take the SPD's election result in 2013 and divide it by the GT proportion of the SPD for the election of 2013 using the same distance d and width w for the respective data window i. Subsequently, the GT prediction of the SPD's 2017 share, is multiplied by the weighting factor *WMC*2 of the SPD. The result is the prediction provided by M2 for the SPD's share in 2017.

$$\text{Weight}_{MC2(p,i,T)} = \frac{\text{Previous election results}_{p,T-4}}{\text{Google proportion}_{i,T-4}} \tag{2}$$

$\text{Weight}_{MC2(p,i,T)}$ partly may account for the possible selection that characterizes individuals that end up in the GT data by including information on shares in the previous election (4 years earlier) in Germany. For instance, non-Internet users are probably underrepresented and younger voters over-represented among GT users. Because the electorate changes for each election, we cannot assume that this weighting method controls/offsets the complete sample bias for the election year we want to predict. But it should provide better predictions than models of class MC1 that are solely based on GT data.[9] In our opinion, this class of models is justified insofar it uses GT and previous election data for the prediction and no other external data.

For models of class MC3, we use polling data from the Infratest Dimap institute for weighting, which is normally published every two weeks.[10] This enables us to calculate a new weighting factor every two weeks, which corrects for possible over- or underestimates in the GT data. This is especially true for short-term trends due to specific events (e.g. television appearances) that lead to an increase in search queries. The weighting factor is calculated similarly to the previous weighting factor, except that in this case we are looking at slices of two weeks of data (the period from one survey to the next).[11] The weighting scheme underlying MC3 models is illustrated in Fig. 3. Figure 3 depicts a data window that contains three election polls, Poll 1–3. Poll 1–3 slice the data window into Time periods 1–4. Moreover, there is polls before the data window namely Poll X1 and X2. We always weight the GT data (proportions) of a time period, e.g., Time Period 2 with a weighting factor. As shown in Eq. 3, the weighting factor consists of the poll at the

---

[9] A special case is MC2 for the 2013 election, for which the 2009 election is used to calculate the weighting factor. The party AfD was founded in 2013, as a consequence, we do not have data from 2009. We circumvented this problem by not weighting the AfD's Google proportion in 2013, which should be taken into account when looking at the results.

[10] Occasionally in 1 or 3 week intervals.

[11] In the actual data the polls do not appear exactly every two weeks.
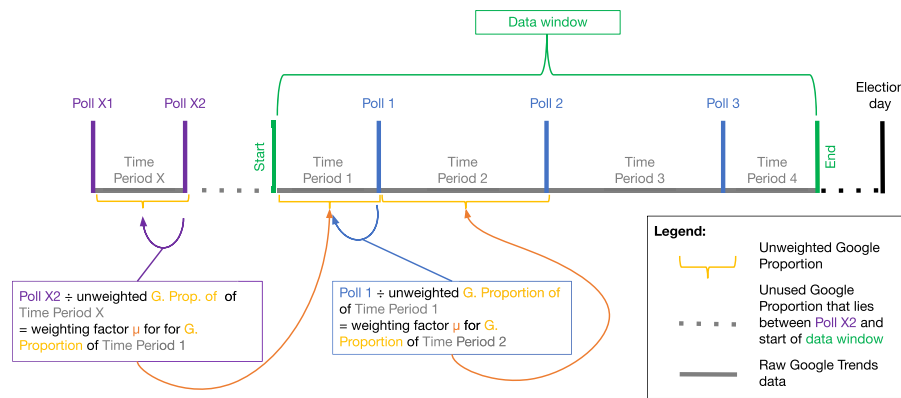
Fig. 3 Weighting logic underlying models of class MC3

start of this time period (here Poll 1) divided by the GT data (proportions) of the previous time period (here Time Period 1). We do the same for the other time periods within the data window, i.e., Time Periods 2 and 3.

$$\text{Weight}_{MC3(t-t+1)} = \frac{\text{Poll}_{\text{Time Period x}}}{\text{Google proportion}_{\text{Time Period x - 1}}} \tag{3}$$

Time Period 1 is special insofar that in most cases the start date of the time window and a poll do not lie on the same date. Therefore the Google proportion between the start of the interval and the 1st poll (here Poll 1) in the time window cannot be weighted. Therefore, we utilize the last poll before the time window (here Poll X2) to calculate the weighting factor. The Google Proportion needed for the weighting factor is thereby limited by the two last polls before the time window (here Poll X1 and X2, resulting in Time Period X).[12]

Results

Comparing predictions across GT data windows, across parties

We start by comparing MC1 models that solely based on GT data, varying the GT data windows both in terms of width, i.e., the number of days covered by the respective GT dataset, and distance, i.e., the number of days the window is away from the election.

In Fig. 4 we visualize a subset of the predictions provided by MC1-GT models for the most recent general election in Germany (26th of September 2021). The most recent election is ideal for benchmarking GT predictions, because usage of the platform has changed over time. Figure 4 plots the prediction errors $(y_p - \hat{y}_p)$ Third, while we argued

---

[12] If there is no poll in our data window, we use the first poll before the data window. In the case where the start time of the data window and a poll fall on the same day, the Google proportion up to the 1st poll before the data window is used and divided by the poll falling on the same day as the start time to calculate the weighting factor. Subsequently, the Google proportion starting one day after the start date and poll up to the first poll in the interval is weighted. Polls that fall on the same day as the end day of an interval are ignored because it wouldn't make sense to calculate a weighting factor using only one day of Google data and applying it on the same day.
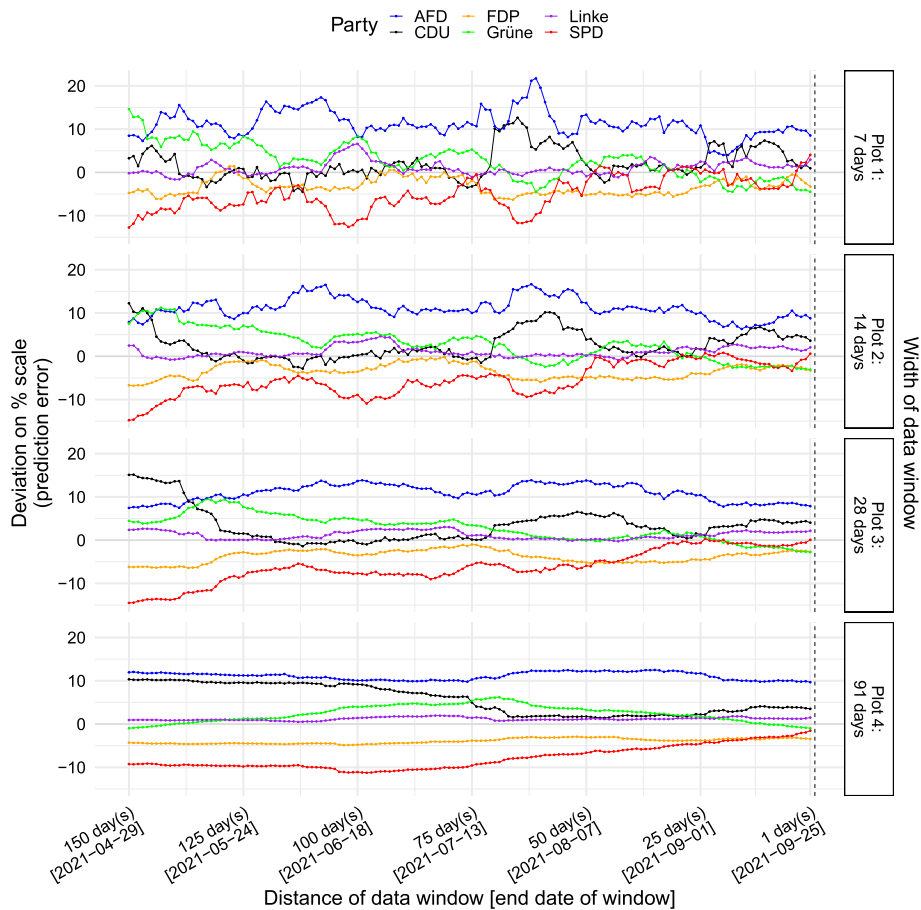
Fig. 4 Accuracy of GT predictions for different parties and party shares across data windows

that as points connected by lines for 3600 predictions = 6 parties * 600 models (defined by 4 widths and 150 distances).[13], [14]

The single panels (Plot 1–4) correspond to the four different widths of the data windows going from 7 days to 91 days (see right-hand y-axis). The x-axis indicates the distance of the respective window to the election. First, we find that the width and distance of the data window to the election matters. We compare models of differing width (from top to bottom) and distance (from left to right). Data windows of low width, e.g., the models based on 7 days of GT data in the top plot, vary much stronger in their predictive accuracy across time. In other words, with such small time spans it matters which week of GT data we have picked for our prediction. Naturally, this variation decreases when we extend the time window on which we base our prediction going from 7 days (Plot 1) to 91 days (Plot 4) at the bottom in Fig. 4. Second, on average accuracy seems to improve slightly the smaller the distance between the GT data

---

[13] Obtained by averaging predictions over 10 GT data samples.

[14] The graph is only showing a subset of our predictions restricted to one election and four data window widths. Overall, from our models that are solely based on GT data we yielded 27,600 predictions, namely predictions for 6 parties of based on 4800 models defined by 8 widths, 150 distances and 4. For the 2009 election, we have 21,600 solely GT data based predictions. For the 2013, 2017, 2021 elections we obtained 21,600 predictions.
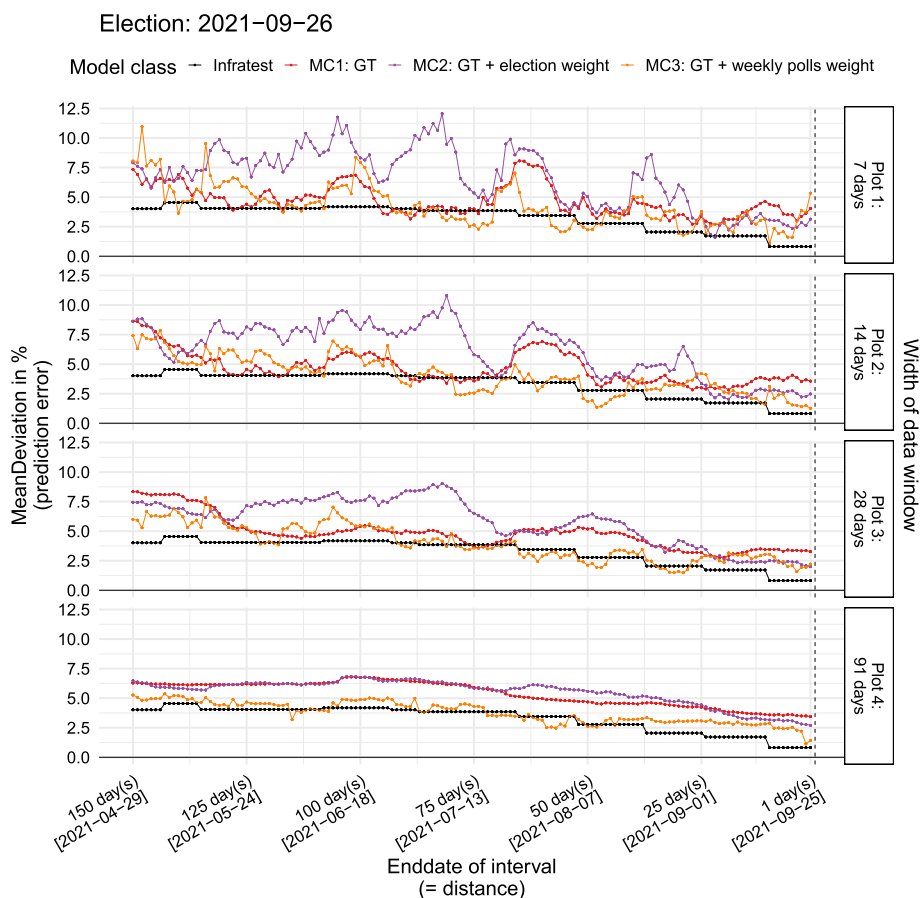
Fig. 5 Comparing the accuracy of different modelling approaches

window and the election. In Fig. 5 we plot the prediction error averaged across all parties for MC1-GT models (red line) contrasting them with other models. Focusing on the red line this analysis also indicates that the error decreases the closer we are to the election. However, we can also clearly see that picking a short data window, e.g., 7 days, results in considerable variation in what regards accuracy (see Fig. 4, Plot 1).

In addition, we used a linear model to model the trend of accuracy as a function of distance holding the width constant at 14 days in Table S2. For the 2021 election, the average error decreases by 0.02% per day distance. In other words, if move the data window 100 days closer to the election, the mean absolute error decreases by 2%. This, however, is not true for the other elections (cf. Table S2). Third, Fig. 4 also highlights the strong variation in accuracy between parties. For instance, GT predictions of the vote share of the Linke are more accurate than, e.g., for the AfD, with the error being closer to 0 across models. For the 2021 election the error generally seems highest for the AfD. Moreover, shares of certain parties are usually underestimated (e.g., SPD), while others are overestimated (e.g., AfD). It seems that the quality of Google searches as a signal of vote choice differs across parties. Importantly, Figure A4 provides further visualizations for more intervals for the 2021 election, Figure A5 provide visualizations for the other elections and Table S2 also provides models for the

Behnert *et al. Journal of Big Data*   (2024) 11:30

Page 16 of 21

other elections. The insights described above are largely confirmed when using data for other elections.

Comparing predictions across model classes

Above, we compared different models within class MC1-GT focusing on the width and distance of data windows. Previous studies have often combined GT data with election or polling data. As described in Section "Prediction models and benchmark" we compare predictions across three broad classes of models. Figure 5 again focuses on the 21' election.

Now, each point represents the mean absolute error, i.e., the prediction error averaged across parties. Figure 5 visualizes 2400 predictions (150 distances * 4 intervals * 4 models). The models are colored: MC1 models based on GT data in red, MC2 models based on GT data weighted with previous election results in purple, MC3 models based on GT data weighted with polling data in orange. Finally, we have the predictions based on Infratest Dimap polls (colored in black), whereby we simply use the last Infratest Dimap poll in the respective time windows.

As in Fig. 4, the four plots in Fig. 5 (Plot 1–4) correspond to different widths of the data window. Not surprisingly, we find that polls (black line) unarguably perform the best for the 2021 election reflected in the fact that prediction error mostly scores below our GT data based models of Class MC1–3. In Fig. 5 this becomes visible as the black line is almost always closer to 0 then the other ones.

We also compared how often our other models actually provide better predictions. GT data based MC1 models provide better prediction in 6% (35 out of 600) of the cases. MC2 models that weight GT data with previous election perform the worst with 0% (2 out of 600) better predictions as compared to polls by Infratest Dimap. Finally, MC3 models that combine GT data with polls perform the best with 24% (147 out of 600) of predictions being better. The picture hardly changes when we restrict our models to those based on a width of 91 days.

Naturally, we could discuss whether the accuracy of our GT based models (MC1) is really so much worse than polling data given the latter's costs. We can collect GT data for free, errors of the size found here may be acceptable for certain applications where we only need rough predictions. Besides, if we want to decrease the dependency of GT-based predictions on short-term trends we should probably base our predictions on a larger time span as shown in Plot in Fig. 5. We provide findings for other elections in Section "Comparing predictions across model classes and elections" and in Additional file 1: Section S4.2.

Comparing predictions across model classes and elections

Above we focused on the most recent election in Germany (26th of September 2021). This focus is justified: Given the changing user population, the most recent election is the most appropriate to assess the predictive power of GT search data. Nonetheless, a comparison with earlier elections can provide us with some insights as to what variations we can expect for coming elections. Figure 6 is similar to Fig. 5 only that the width of the data window is held constant at 91 days which corresponds to Plot 4 in Fig. 5, and predictions for different elections are shown on top of each other (Plot
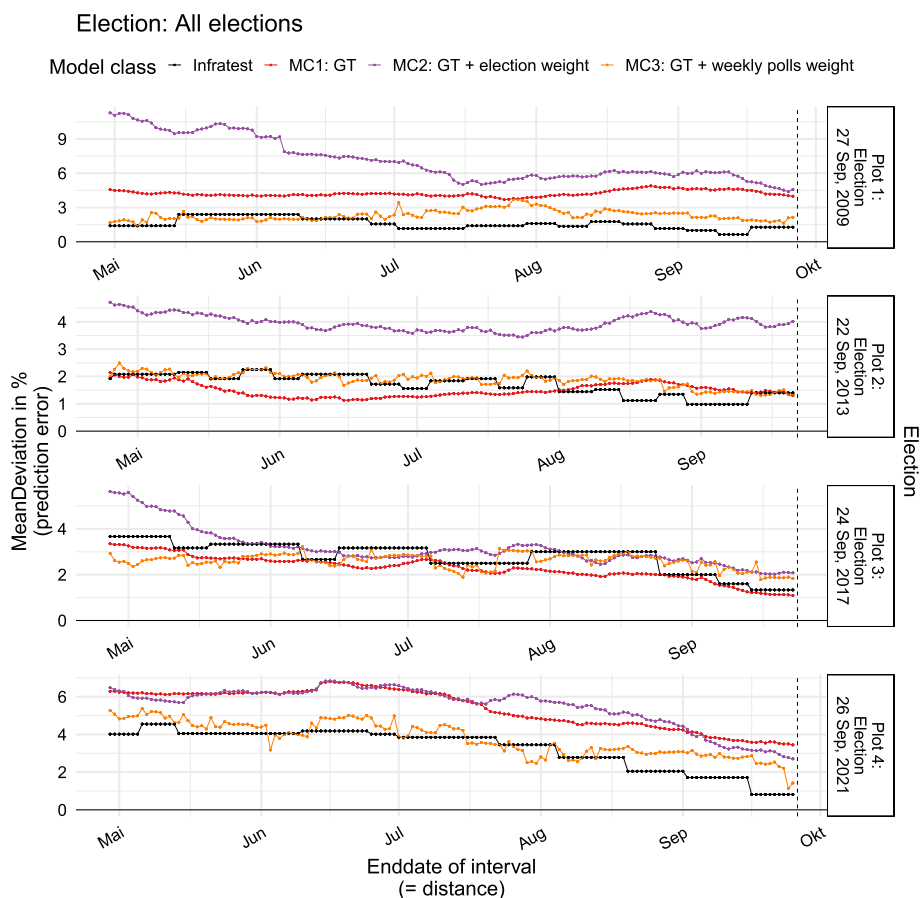
Fig. 6 Accuracy of predictions across model classes and election

1–4). Again, each point represents the mean absolute error, i.e., the prediction error averaged across parties whereby predictions by different model classes are colored.

We find that indeed the accuracy of predictions varies across elections. Table 3 provides an overview regarding how often the different models where better than purely poll based predictions across elections (poll based predictions always concern the last poll). We can see that 2021 and 2009 were particular bad years for models of class MC1 and MC2 with zero better predictions than polls. In contrast, models of class MC3 provided better predictions for 20% of the estimated models both in 2009 and 2021. For the 2013 election, the purely GT data based models beat the polls in 66% of the estimated models and provide better predictions than the other model classes most of the time. For the 2017 election, the purely GT data based models fare even better, beating the polls in 97% of the models (defined by width and distance), with the other two model classes faring somewhat worse. Hence, in general, 2017 was a particularly bad year for the polls in our data.

We can only speculate why there is such strong variation across elections. First, as mentioned, we have to assume that the representativity of Google search users varies across time. In principle, it is possible that the population of Google search users was more representative of German voters in the 2013, 2017 elections than in 2009 or

Table 3 Comparing model predictions to polls

Model class	Election	Better predictions: %	Better predictions: n out of total
MC1: GT	2009	0	0 out of 150
	2013	66	99 out of 150
	2017	97	145 out of 150
	2021	0	0 out of 150
MC2: GT + election weight	2009	0	0 out of 150
	2013	0	0 out of 150
	2017	32	48 out of 150
	2021	0	0 out of 150
MC3: GT + weekly polls weight	2009	20	30 out of 150
	2013	29	43 out of 150
	2017	73	109 out of 150
	2021	20	30 out of 150

2021. Second, it could be related to the elections themselves. Potentially, the extend to which the action "search for" is related to "vote for" depends on the election with more "charged" elections like the 2021 election decreasing this correlation [31]. Third, it is possible that the accuracy of the polling data which we use as a benchmark varies over time.

Conclusion

Google trends data has become a popular data source for prediction in different domains such as elections. While most previous research focuses on binary electoral outcomes such as referendum results or presidential elections, we evaluate GT predictions in a multi-party setting. We developed a framework that allows us to compare predictions across 1200 fine-grained GT data windows that vary both in terms of width (7 to 91 days) and distance to the election (1 to 150 days). And, we compare predictions across several elections (four general elections in Germany). Besides, we provide a more systematic assessment hitherto neglected choices such as the selection of search terms, GT data samples and search refinement categories. Tackling these different dimensions allows us to provide some unique insights. First, we find that predictive accuracy varies significantly as a function of the width and distance of GT data windows. To some extent disagreement on the predictive accuracy of GT data in previous research may be linked to the varying choices researchers made here (cf. Table 2). In what concerns width(s), accuracy varies significantly across time for shorter data windows (here 7–28 days). We would generally recommend choosing larger data windows for predictive purposes to average out variation that is due to singular events. And, akin to classic polling data we find that predictive accuracy increases the closer our data window to the election. Second, in terms of predictive models we compared models based on GT data, on GT + previous electoral outcomes data and GT + opinion poll data. Generally, high quality opinion polls, in our case by the renowned company Infratest Dimap, still represents the benchmark in terms of accuracy, especially in the case of the latest election in 2021. We found that models that combine polling data with GT data (MC3) data fare better than

purely GT-data-based models (MC1). Furthermore, we find that weighting GT data with previous electoral outcomes [14] does not help predictive accuracy, at least not for the German case (cf. Fig. 5). Third, while we argued that recent elections are the appropriate benchmarks for GT data predictions, comparisons to past elections may still be revealing. Especially, insofar our findings for the 2021 election are somewhat discouraging, which goes against earlier studies that emphasize the predictive accuracy of GT data. Hence, we need to test whether our findings for 2021 also hold for other elections. And, especially for the 2013 and 2017 election GT data models where much more accurate, at times beating or at least aligning with predictions based on opinion polls (cf. Fig. 6). Above we speculated that the nature of the election may affect the predictive power of GT data (cf. [31]). More generally, this highlights that conclusions regarding predictive accuracy from one election may not generalize to other elections. Finally, we conclude that GT prediction research should ideally become more transparent. From our review we learned that identifying the various characteristics in Table 1, i.e., all representing important choices when using GT data, was a challenge. Future research would benefit if authors would report more details, e.g., the exact time of GT data collection, the number of GT samples, the nature of search term selection. For instance, we found that it matters which category search filter is selected (see Additional file 1: Section S1).

Our study has several limitations that provide venues for future research (see [32] for a systematic discussion of limitations and a literature review). First, we pursued a systematic approach to search term selection, testing different search terms against each other (see Section "Search terms and category filter"). Future research might benefit from choosing an even wider net of search terms, trying out different ones. Second, while we predicted elections shares we did not forecast them in that the elections happened already. Since, we didn't built any sophisticated models we somewhat circumvented the danger of adapting our models to already seen data. Nonetheless, as is now more common in the literature on election forecasting [33], studies that rely on GT data for prediction could also be pre-registered. Third, from the general perspective of election forecasting, GT data could be seen as just another dynamic signal that can be integrated into more sophisticated modelling strategies (e.g., [4, 34]) that go beyond the weighting approach used in the present paper. Whether such additional sources of data can improve predictions remains to be tested. Fourth, the changing predictive power of GT across elections may be related to (1) how representative the GT data is of searches on the platform and (2) how representative users of the GT platform are of voting the population. Pertaining to (1), both aspects of representiveness should be studied further. Finding variation across GT data samples Raubenheimer et al. [21] suggested that there might be a publication bias, i.e., successful predictions are based on suitable samples. For this reason we averaged across GT data samples (cf. Section "Collecting GT data" and Additional file 1: Section S2). However, more research into the variability of such samples for particular predictions is warranted. And, user representativity (2) is relevant as well. Platforms are generally quite secretive about their userbase. Collecting more evidence on how representative Google users are of the general (voting) population would help to explain the accuracy of corresponding predictions and help to develop more sophisticated weighting schemes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40537-023-00868-4.

Additional file 1: Figure S1. Category selection among supercategories of Google Trends. **Figure S2.** Search terms for other parties. **Figure S3.** Sampling error variation for the 2021 election. **Figure S4.** Accuracy of GT predictions for different parties and party shares across datawindows (more intervals). **Figure S5.** Accuracy of GT predictions for different parties and party shares across datawindows (more intervals, elections). **Figure S6.** Accuracy of predictions across model classes for 2009 election. **Figure S7.** Accuracy of predictions across model classes for 2013 election. **Figure S8.** Accuracy of predictions across model classes for 2017 election. **Figure S9.** Comparison across elections. **Figure S10.** Benchmarking against other polls (Elections: 2009, 2013). **Figure S11.** Benchmarking against other polls (Elections: 2017, 2021). **Figure S12.** Figure 4 with 'other parties. **Figure S13.** Figure 5 with 'other parties' (2021 election). **Figure S14.** Figure 5 with 'other parties' (2009 election). **Figure S15.** Figure 5 with 'other parties' (2013 election). **Figure S16.** Figure 5 with 'other parties' (2017 election). **Figure S17.** Figure 6 with 'other parties. **Table S1.** Overview datasets across samples that are different. **Table S2.** Prediction accuracy as a function of distance days.

Availability of data and materials
Code and data to reproduce the study are available in the Additional file 1: https://doi.org/10.7910/DVN/1EUPZU.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
The authors declare consent to publish.

Competing interests
The authors declare that they have no competing interests in relation to this research, whether fnancial, personal, authorship or otherwise, that could afect the research and its results presented in this paper

References
1. Choi H, Varian H. Predicting the present with Google trends. Economic Record. 2012;88:2–9.
2. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009;457(7232):1012–4.
3. Prado-Román C, Gómez-Martínez R, Orden-Cruz C. Google trends as a predictor of presidential elections: The United States versus Canada. Am Behav Sci. 2021;65(4):666–80.
4. Stoetzer LF, Neunhoeffer M, Gschwend T, Munzert S, Sternberg S. Forecasting elections in multiparty systems: a Bayesian approach combining polls and fundamentals. Polit Anal. 2019;27(2):255–62.
5. Wang W, Rothschild D, Goel S, Gelman A. Forecasting elections with non-representative polls. Int J Forecast. 2015;31(3):980–91.
6. Granka L. Using online search traffic to predict US presidential elections. PS Polit Sci Polit. 2013;46(2):271–9.
7. D21 I. Share of internet users in Germany from 2001 to 2021 Graph]. Statista. 2022. https://www.statista.com/statistics/380514/internet-usage-rate-germany/
8. StatCounter. Desktop and mobile search market share of search engines in Germany in December 2023 [Graph]. Statista. 2023. https://www.statista.com/statistics/445974/search-engines-market-share-of-desktop-and-mobile-search-germany/
9. Jun SP, Yoo HS, Choi S. Ten years of research change using Google trends: from the perspective of big data utilizations and applications. Technol Forecast Soc Change. 2018;130:69–87.
10. Kandula S, Shaman J. Reappraising the utility of Google Flu trends. PLoS Computat Biol. 2019;15(8): e1007258. https://doi.org/10.1371/journal.pcbi.1007258.
11. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. Science. 2014;343:1203–5. https://doi.org/10.1126/science.1248506.
12. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. Proc Natl Acad Sci. 2015;112(47):14473–8.
13. Brodeur A, Clark AE, Fleche S, Powdthavee N. COVID-19, lockdowns and well-being: evidence from Google trends. J Publ Econom. 2021;193: 104346.

14. Polykalas SE, Prezerakos GN, Konidaris A. A General Purpose Model for Future Prediction Based on Web Search Data: Predicting Greek and Spanish Election. In: 2013 27th International Conference on Advanced Information Networking and Applications Workshops. 2013. p. 213–8. https://doi.org/10.1109/WAINA.2013.155

15. Wolf JT. Trending in the right direction: using Google trends data as a measure of public opinion during a presidential election. Blacksburg: Virginia Tech; 2018.

16. Askitas N. Calling the Greek Referendum on the Nose with Google Trends. SSRN Electron J. 2015. https://doi.org/10.2139%2Fssrn.2708382

17. Askitas N. Predicting the Irish 'Gay Marriage' Referendum. SSRN 2674243. 2015.

18. Mavragani A, Tsagarakis KP. Predicting referendum results in the big data Era. J Big Data. 2019;6(1):1–20.

19. Polykalas SE, Prezerakos GN, Konidaris A. An algorithm based on Google Trends' data for future prediction. Case study: German elections. IEEE International Symposium on Signal Processing and Information Technology; 2013. p. 000069–73.

20. Sjövill R. Using Search Query Data to Predict the General Election: Can Google trends help predict the Swedish General Election? 2020.

21. Raubenheimer JE, Riordan BC, Merrill JE, Winter T, Ward RM, Scarf D, et al. Hey Google! Will New Zealand vote to legalise cannabis? Using Google trends data to predict the outcome of the 2020 New Zealand cannabis referendum. Int J Drug Policy. 2021;90: 103083.

22. Raubenheimer JE. A practical algorithm for extracting multiple data samples from google trends extended for health. Am J Epidemiol. 2022;191(9):1666–9.

23. Mavragani A, Tsagarakis KP. YES or NO: Predicting the 2015 GReferendum results using Google Trends. Technological Forecasting and Social Change. 2016;109:1–5. https://www.sciencedirect.com/science/article/pii/S0040162516300580

24. Harkan AA. Predicting the results of the 2019 Indonesian presidential election with Google trends Asia-Pacific research in social sciences and humanities Universitas Indonesia (APRISH 2019). Amsterdam: Atlantis Press; 2021.

25. Yasseri T, Bright J. Can electoral popularity be predicted using socially generated big data? Abstract it - Information Technology. 2014;56(5):246–253. https://doi.org/10.1515/itit-2014-1046

26. Vergara-Perucich, F. Assessing the accuracy of Google trends for predicting presidential elections: The case of Chile 2006–2021. Data. 2022;7(11):143. https://doi.org/10.3390/data7110143

27. Help GT. FAQ about Google Trends data - Trends Help. 2023. https://support.google.com/trends/answer/4365533?hl=en &amp%3Bref_topic=6248052

28. Help GT. Refine Trends results by category - Trends Help. 2023. https://support.google.com/trends/answer/4359597?hl=en &ref_topic=4365530

29. Eichenauer VZ, Indergand R, Martínez IZ, Sax C. Obtaining consistent time series from Google Trends. Econom Inquiry. 2022;60(2):694–705.

30. Stephens-Davidowitz S, Varian H. A hands-on guide to google data. further details on the construction can be found on the Google Trends page. 2014;

31. Faas T, Klingelhöfer T. German politics at the traffic light: new beginnings in the election of 2021. West Eur Polit. 2022;45(7):1506–21. https://doi.org/10.1080/01402382.2022.2045783.

32. Mehltretter J, Keusch F, and Sajons C. The (mis)use of Google Trends data in the social sciences - A systematic review, critique, and guidelines. Working paper. 2023;1–55. https://doi.org/10.17605/OSF.IO/CTN63

33. Gschwend T, Müller K, Munzert S, Neunhoeffer M, Stoetzer LF. The Zweitstimme model: a dynamic forecast of the 2021 German federal election. PS Polit Sci Polit. 2022;55(1):85–90.

34. Munzert S, Stötzer L, Gschwend T, Neunhoeffer M, Sternberg S. Zweitstimme. Org. Ein strukturell-dynamisches Vorhersagemodell für Bundestagswahlen. Politische Vierteljahresschrift. 2017;418–41.

Publisher's Note