This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TCSS.2021.3063660

A Systematic Review of Predicting Elections Based on Social Media Data: Research Challenges and Future Directions

Kellyton dos Santos Brito, *Member, IEEE*, Rogério Luiz Cardoso Silva Filho, and Paulo Jorge Leitão Adeodato

***Abstract*—The way politicians communicate with the electorate and run electoral campaigns was reshaped by the emergence and popularization of contemporary social media (SM), such as Facebook, Twitter, and Instagram social networks (SN). Due to inherent capabilities of SM, such as the large amount of available data accessed in real time, a new research subject has emerged, focusing on using SM data to predict election outcomes. Despite many studies conducted in the last decade, results are very controversial, and many times challenged. In this context, this work aims to investigate and summarize how research on predicting elections based on SM data has evolved since its beginning, to outline the state of both the art and the practice, and to identify research opportunities within this field. In terms of method, we performed a systematic literature review analyzing the quantity and quality of publications, the electoral context of studies, the main approaches to and characteristics of the successful studies, as well as their main strengths and challenges, and compared our results with previous reviews. We identified and analyzed 83 relevant studies, and the challenges were identified in many areas such as process, sampling, modeling, performance evaluation and scientific rigor. Main findings include the low success of the most-used approach, namely volume and sentiment analysis on Twitter, and the better results with new approaches, such as regression methods trained with traditional polls. Finally, a vision of future research on integrating advances on process definitions, modeling, and evaluation is also discussed, pointing out, among others, the need for better investigating the application of state-of-art machine learning approaches.**

***Index Terms*—Elections, Social Media, Social Networks, Machine Learning, Systematic Review**

I. Introduction

Social media (SM) has played a central role in politics and elections throughout this decade. We have entered a new era mediated by SM in which politicians conduct permanent campaigns without geographic or time constraints, and extra information about them can be obtained not only by the press, but also directly from their profiles on social networks (SNs) and through other people sharing and amplifying their voices on SM. In this new scenario, SM is used extensively in electoral campaigns [1], and an online campaign's success can even decide elections. In practice, recent examples of SM engagement and electoral success include the 2016 U.S. presidential election, when Donald Trump focused his campaign on free-media marketing [2], and the 2018 Brazilian presidential election, when the candidate with more SM engagement but little exposition on traditional media was elected [3].

Moreover, in some way, it is possible to measure how a politician's message is spreading over SM and try to estimate how much attention a candidate is receiving or how many people are talking about a candidate. Thus, considering the large amount of data available in real time and the low cost of their acquisition, combined with the advances of techniques for processing them, a new research subject has emerged, focusing on using SM data to predict elections outcomes.

Only two years after Twitter and Facebook's launch for general public, studies with the objective of predicting elections based on SM data started to be published: Tilton [4] can be considered a preliminary study focused on student elections, published in 2008. Also, two studies published in 2010 at the same forum, Tumasjan *et al.* [5] and O'Connor [6], are considered seminal studies regarding predicting political elections based on SM. The former presented an approach based on volume counting of posts on Twitter (tweets), and the latter was based on the sentiment extracted from those tweets.

One decade after Tumasjan's and O'Connor's seminal studies had claimed promising results, several initiatives focused on predicting elections all around the world, such as in Europe [7], [8], Asia [9], [10], Latin America [11], [12], Africa [13], [14] and U.S. [15]–[17], just to cite some. These studies presented a variety of methods, were applied in many different electoral scenarios, used different SNs as information source, and had different outcomes. In fact, many studies claimed very positive results, others challenged the predictive power of SM, and even the same study may achieve positive

Manuscript received June 01, 2020; revised January 08, 2021; accepted February 20, 2021. Date of publication XXXX. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Bin Hu.

Kellyton dos Santos Brito is with the Centro de Informática, Universidade Federal de Pernambuco and Departamento de Computação, Universidade Federal Rural de Pernambuco, 52171-900 Recife/PE, Brazil (e-mail: kellyton@kellyton.com.br).

Rogério Luiz Cardoso Silva Filho and Paulo Jorge Leitão Adeodato are with the Centro de Informática, Universidade Federal de Pernambuco, 50740-560 Recife/PE, Brazil (e-mail: rlcsf@cin.ufpe.br@cin.ufpe.br, pjla@cin.ufpe.br).

results in one context and negative in another [18].

Thus, there is not yet a common perspective on the literature or well-established methods, processes, and tools for predicting elections results based on SM data. Moreover, even the SM context has changed over the years. For example, Facebook surpassed the number of active users of Twitter, and also new SN has emerged, such as Instagram.

In this context, this work aims to give a thorough review and investigation of the state of both the art and practice of predicting election outcomes based on SM data and identify key research challenges and opportunities in this field. We systematically review 83 studies from 2008 to 2019, identify the context of studies, main models, strengths, and challenges of this new area, as well as the main characteristics present on successful studies, and present a deep discussion about future directions.

The remainder of this paper is organized as follows: Section II presents the background and previous studies related to this work, as well as an analysis of the main points of similar comparative studies. In Section III we present the review method and procedure employed in this study, followed by Section IV, which provides an overall summary of the selected studies and assess their quality. In Section V we discuss the answers to three of the predefined research questions regarding the electoral context of studies, main approaches, and main characteristics of successful studies. Section VI answers the last research question regarding main strengths and challenges, summarizing the results, and ends with a discussion about future directions. Section VII presents a comparison with previous works and reviews the limitations of this study, followed by Section VIII, that concludes and summarizes the outcomes.

II. Research Background

A. The rise of elections prediction with SM data

Contemporary SM systems are new: Facebook and Twitter were launched to the public in 2006, and Instagram emerged in 2010. The use of SM in modern political activities and being considered a source for election prediction started just a little after their launch.

One of the first attempts that aimed at predicting election outcomes using data from SM may be attributed to Tilton [4]. In 2008, only two years after Facebook's launch for the general public, he tried to predict election outcomes of a connected society, in this case a university, framed by the following research question: "Could Facebook be used to estimate the results of a student election?" Results showed his model was able to predict what place the candidates came in 21 out of 27 times in a given election. Probably because it is not related to formal politics scenario, Tilton's study is seldom cited by studies in the area, but we consider it as a very insightful preliminary study in this field.

Two studies can be considered seminal in predicting political elections with SM data and are cited by almost all following studies. In 2010, Tumasjan *et al.* [5] presented a study on the 2009 German federal election. They collected all the tweets with the names of any of six parties represented in the German parliament, or prominent politicians of these parties, and compared the volume of tweets with the election results. Per their results, they claimed that "the mere number of tweets mentioning a political party can be considered a plausible reflection of the vote share and its predictive power even comes close to traditional election polls." In the same year and with an approach improved by a sentiment detection of tweets, O'Connor [6] found that "a relatively simple sentiment detector based on Twitter data replicates consumer confidence and presidential job approval polls."

Based on these two studies, the volume of tweets combined with automatic sentiment detection became the main approach for most further research around the world, such as in the Netherlands [7], Italy and France [8], India [9], Indonesia [10], Colombia [11], Chile [12], and the U.S. [15]. In general terms, researchers collected tweets referring to a candidate or party; performed a sentiment analysis to classify the post as positive, negative, or neutral; and tried to correlate the volume of positive and negative posts with electoral results. In these studies, the main challenges were gathering data via an open search on Twitter and the sentiment analysis.

Despite being the most-used approach, the analysis of the volume and sentiment of tweets engendered a number of criticisms just after their launch [19]–[21]. In fact, by using these approaches, results can vary widely, as discussed by Jungherr [22]. After replicating Tumasjan's seminal study, Jungherr argued that "the results are contingent on arbitrary choices of the authors," and indicated that simply including one more party or day of collection would greatly change the results.

Moreover, despite criticism, recent works still used similar approaches to the volume and/or sentiment of tweets and achieved a variety results, both positive [23], [24], negative [16], [25] and even mixed results [15], [18]. Additionally, novel approaches started to appear, such as models based on regression or time series methods [26], [27], and models using traditional polls for training or comparing results to calibrate the model [28].

B. Analysis of previous reviews

Due to the variety of approaches, with different achieved results even in replications of the same approach in the same context [22], some researchers tried to summarize the knowledge in this area.

In 2013, Kalampokis *et al.* [29] presented a systematic review aiming to understand the predictive power of SM, not only on the electoral context. By analyzing 52 studies, 11 regarding election predictions, they identified that main approaches were based on volume, sentiment, and user profiling. In addition, the use of predictive analysis using linear regression was identified, but not on the studies related to the political context. Also, they verified that 40% of studies that had used sentiment-related variables challenged SM predictive power, i.e., was not successful, and this number increased to 65% in the case of lexicon-based approaches. Finally, they emphasized the lack of predictive analytics evaluation and controversial results of electoral predicting studies.

In the same year, Gayo-Avello [30] presented a study that we consider the first review specifically on predicting elections with SM, focused on Twitter. By analyzing 10 previous studies from 2010 to 2013, he concluded that "the presumed predictive power regarding electoral prediction has been somewhat exaggerated." Moreover, as in [29], he identified volume and sentiment analysis as main approaches and the need to use more up-to-date methods for sentiment analysis. Also, he expanded the list of challenges, such as the dependency of arbitrary decisions made by researchers regarding keywords, parties, candidates and selection of the data collection period, and problems related to Twitter, such as demographic and self-selection bias, and bias related to spam, misleading propaganda and astroturfing. He ended the study pointing out that regression models may be a future direction.

In 2015, studies from Prada [31] and O'Leary [32] presented in general lines the main approaches for predicting using Twitter in many different domains, and briefly described a few studies related to election predictions (2 and 11 studies respectively). In 2018, Kwak [33] presented results of a survey including 69 papers which supported the argument that SM can be used in understanding political agenda, rather than in election forecast. Ultimately, most recent studies [34][35] presented limited nonsystematic surveys, both analyzing 13 papers, adding some arguments to the original review from Gayo-Avello [30]. Koli [34] argued that prediction using Twitter can have better results in developed countries, due to a higher literacy rate and internet access, than in developing countries. In addition, Bilal [35] considered the challenges of sentiment analysis in languages other than English. Despite these new arguments, recent studies fail to identify novel approaches, as well as approaches using SM other than Twitter and Facebook.

There is not yet a common consensus on the literature regarding well-established methods, processes, and tools for predicting election results based on SM data. Moreover, the SM landscape is undergoing continuous changes, as well as patterns of use. For example, Facebook surpassed the number of active users of Twitter, and even new SNs have become more popular, such as Instagram [36]. Thus, a thorough review providing an understanding of the past and directions for future research is still needed and should be updated frequently until common bases can be defined.

III. Methodology

The method chosen for this research was a systematic literature review, which has proven to be a replicable and effective manner with which to identify, evaluate, interpret and compare studies that are relevant to a particular question or area [37]–[40]. The method used in this research follows the guidelines defined by [40] and is fully described in Appendix I. This section presents the main points.

A. Research Questions

To define the research questions of this study, we returned to the main objectives:

To provide a thorough review and investigation of the state of both the art and the practice of predicting election outcomes based on SM data and to identify key research challenges and opportunities in this field.

Then, the following research questions were derived:

- RQ1: In which electoral contexts is the research being performed?

This question aims at identifying the electoral contexts being studied, such as the year and country in which the election took place, and the type of election. This question is intended to ascertain whether the studies are best suited or giving attention to any particular electoral context.

- RQ2: What are the main approaches?

The objective of this question is to identify the main approaches used, their main characteristics, how they are modeled and applied to predict elections, and which are the metrics used to assess their performance.

- RQ3: What are the main characteristics of successful studies?

The objective of this question is to identify the main characteristics of allegedly successful studies, in order to identify in which specific contexts, which approaches, and which factors yield effective results.

- RQ4. What are the main strengths and challenges of predicting elections with social media?

After studying the context, approaches and characteristics of successful studies, the answer to this question aims to summarize the main perceived strengths, weaknesses, challenges, and opportunities in this new research area to guide future research.

B. Search Process

The rigor of the search process is one of the distinctive characteristics of systematic reviews [38]. To implement an unbiased and strict search, two approaches were combined: (i) automated search on indexing systems and (ii) snowballing search on the references of studies found on the automated search.

The automated search was performed in four indexing systems: ACM Digital Library, IEEEXplore Digital Library, ISI Web of Science, and Scopus. The search was performed on papers' metadata: title, abstract, and keywords and aimed to find studies focused on predicting elections based on SM data. Then, after some initial refinements, the following search string was used in the automatic search:

(model OR method OR approach OR framework) AND (predict) AND (election*) AND ("social media" OR twitter OR facebook OR instagram)*.

The snowballing search on the references was applied only at the end of study selection to perform this search only on already identified relevant studies.

C. Quality Assessment

One initial difficulty regarding the quality assessment is that there is no established manner with which to define study "quality." In this study, we used the premise suggested by [37], in which quality relates to the extent to which the study minimizes bias and maximizes internal and external validity. Thus, we focused the quality assessment on the rigor of the study. Hence, we proposed the following quality assessment questions:

- QA1: Are the aim(s)/objective(s) clearly identified?
- QA2: Is the related work comprehensively reviewed?
- QA3: Are the findings/results clearly reported?
- QA4: Are bias and threats to validity clearly discussed?
- QA5: Did the study compare the proposed solution and results with other works?

IV. Review Results

In this section, we provide the results of study selection, an overall summary of the selected papers, and the quality assessment result. The findings and answers to the predefined research questions are discussed in following sections.

A. Study Selection

The search procedure was performed twice. The first execution was completed on July 31, 2019 and generated the first version of this study with incomplete data about the 2019 year. One year later, a new search was performed to gather the remaining papers published in 2019.

After the phases described in Appendices I and II, the study selection resulted in a final set of 90 studies: 83 main primary studies and 7 surveys or literature reviews. Primary studies will be analyzed and discussed to answer the research questions, whilst surveys will be used in the discussion and comparison of this works' results. The list of full references and the summary of collected data of all 83 primary studies is presented in Appendices III and IV.

B. Overview of Selected Studies

The 83 selected studies were conducted by a total of 224 authors and co-authors from 105 institutions in 28 countries. Most authors (194 – 87%) were involved in only one study, 29 authors (13%) were involved in two studies, and only one author, Daniel Gayo-Avello, was involved in three studies. In the same way, the majority of institutions (93 institutions – 89%) were also involved in only one study, 10 institutions (10%) were involved in two studies, and only the Universidad de Oviedo, Gayo-Avello's institution, and Università degli Studi di Milano were involved in three studies. Moreover, research in this subject was spread out among 28 countries, with focus on the U.S. (19 institutions), India (11), Indonesia and the U.K. (8), China and Italy (6), and Germany and Taiwan (5). Figure 1 presents the geographical distribution of the research among countries, and Table I presents the list of countries, number of institutions, and studies published by these institutions. The sum may naturally be different from the total number of studies due to multiple relations between authors, institutions, and studies.

These data show that the research in this area is performed by institutions in all continents, being only two institutions in Africa, and the U.S. being the country with more institutions involved. Nevertheless, we did not find prominent researchers, research groups, or clusters performing a sustainable research in the area. In fact, even the work of Gayo-Avello was performed in a limited period, mostly in 2011 and 2012, and studies from Università degli Studi di Milano was limited to 2013 and 2014. The absence of well-established research clusters is not surprising because modern SM are recent and its political use is also a new phenomenon. Thus, the research in this area is still in the beginning.

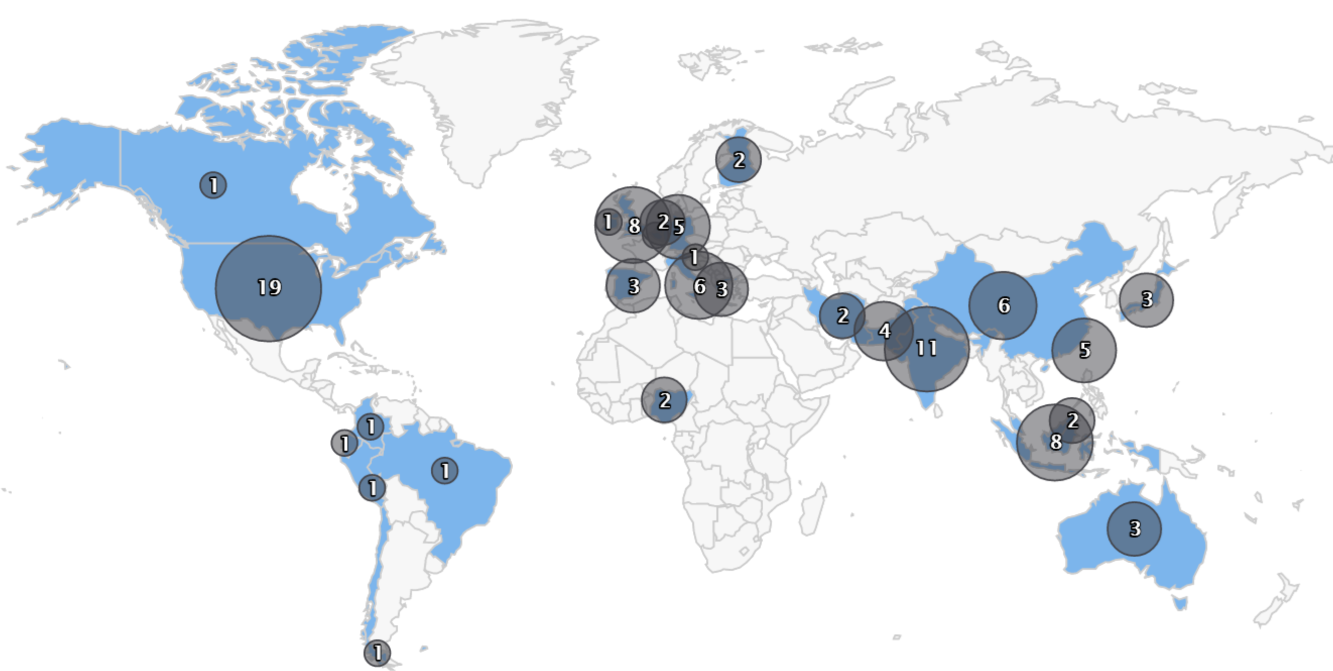

Fig. 1. Institutions distribution among countries

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TCSS.2021.3063660

TABLE I. NUMBER OF INSTITUTIONS AND STUDIES BY COUNTRY

Country	Institution	Studies
United States	19	18
India	11	13
United Kingdom	8	7
Indonesia	8	6
China, Italy	6	5
Germany	5	4
Taiwan	5	3
Pakistan	4	4
Spain	3	5
Australia, Greece	3	3
Japan	3	2
Finland, Hong Kong, Iran, The Netherlands	2	2
Malaysia, Nigeria	2	1
Chile, Ecuador	1	2
Belgium, Brazil, Canada, Colombia, Ireland, Peru, Slovenia	1	1

These 83 studies scattered amongst 72 forums, 53 (64%) were published in conferences or workshops proceedings, and 30 (36%) were published in academic journals. Due to the exclusion of short papers, selected papers had a mean of 10 pages in length and a median of 8 pages. The forum that published more studies (4), is the *International AAAI Conference on Weblogs and Social Media*, followed by the *IEEE International Conference on Data Mining Workshops* (3 studies). Only two journals published more than one study: The *Social Network Analysis and Mining*, and the *Social Science Computer Review* published two studies. Also, 64 forums published only one study. These data reveal that there is not yet a common well-known forum for publication on this subject.

Regarding publishing years, after Tilton's preliminary study in 2008 [4] the studies by Tumasjan [5] and O'Connor [6], in 2010, were considered the seminal papers in this area, being cited by almost all following papers. After these, as shown in Fig. 2., the interest in this subject slowly increased until 2014, and increased markedly starting in 2015. This behavior can be explained by the intense use of SM, especially Twitter, during Trump's campaign for the 2016 U.S. presidential elections [2], and for the same reason an increase of the number of studies after 2020 U.S. presidential elections is expected.

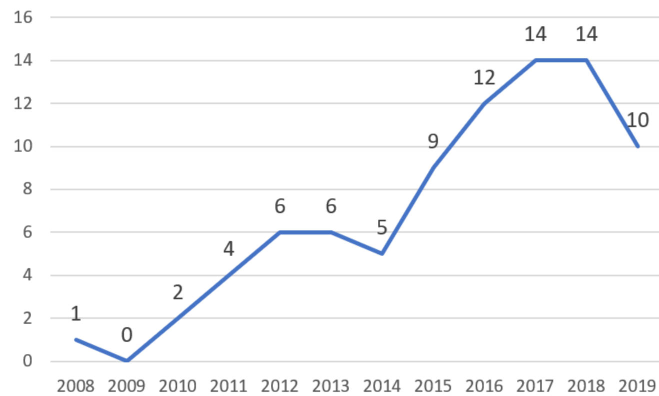

Fig. 2. Study distribution over the publication years.

C. Quality Assessment

As for the quality assessment, the objective was not to exclude any study based on measured quality, rather to understand the general quality of published studies, and to detect possible strengths or weaknesses on methodology. Thus, instead of listing the detailed quality scores, we present in Table II the distribution of the studies over the quality assessment questions and highlight the main results.

TABLE II. DISTRIBUTION OF STUDIES OVER QUALITY

Quality Assessment Question	Yes	Partially	No
QA1: Are the aim(s)/objective(s) clearly identified?	98,8%	1,2%	0%
QA2: Is the related work comprehensively reviewed?	89,2%	4,8%	6,0%
QA3: Are the findings/results clearly reported?	94,0%	6,0%	0%
QA4: Are bias and threats to validity clearly discussed?	44,6%	16,9%	38,6%
QA5: Did the study compare the proposed solution and results with other works?	10,8%	51,8%	37,3%

Considering "QA1: Are the aim(s)/objective(s) clearly identified?," almost all studies were clear on their objectives, but only one [PS15] was unclear on whether the main objective was to predict elections or just to compare sentiment analysis algorithms.

Regarding "QA2: Is the related work comprehensively reviewed?," most studies (89%) presented a comprehensive review of related works, four studies (4.8%) made a very brief presentation, and five studies (6%) presented no works reviewed that were related to election predictions (except one of these that presented a review on sentiment analysis). Considering "QA3: Are the findings/results clearly reported?," almost all studies (94%) clearly presented their results, and five studies (6%) were not noticeably clear in results presentation. Studies with no clear results usually report the results of sentiment analysis and fail to directly correlate it with the prediction of election results.

The main concerns reside in RQ4 and RQ5. Considering the discussion of the study's bias and threats to validity, only 37 studies (45%) detailed discussion, whereas 32 studies (39%) presented no discussion at all, and 14 studies (17%) presented brief discussions, mainly for justifying negative results.

Finally, the analysis of question "QA5: Did the study compare the proposed solution and results with other works?" shows an important deficiency of many studies in this subject. Only nine studies (11%) performed a clear comparison and discussion of their results with other research, whereas more than a third (31 studies – 37%) did not compare results at all. The remaining 52% studies were classified as "partially" in comparison. In 11 studies (13% of the total) they performed inner comparisons, that is, they implemented other approaches (mainly Tumasjan's [5] volume counting) and compared results, whereas 32 studies (39% of the total) compared their results with traditional polls.

These data lead to the first conclusion of this study: many studies claim positive (or negative) results, but it is hard to

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TCSS.2021.3063660

support those results because no comparison against previous research has been carried out.

V. Discussion of Research Questions

In this section, we address three of the four research questions presented in Section III.

A. RQ1: In which electoral contexts is the research being performed?

For contextual definition, we first studied the number of elections characterized by the study. We grouped studied elections in three sets: (a) unique elections, in which the studied cases consisted of only a single election, occurring in a specific year, for one position, with a limited set of candidates; (b) one election with sub-elections, in which the studied cases consisted of parallel elections for different positions, such as a mayoral election in three different cities of the same country at same time, or senate elections in different states in the same electoral context; and (c) many elections, in which the cases studied consisted of elections in totally different contexts, such as different countries or different years. Table III shows that most studies were performed with only one election (60 studies – 72%), 12 studies (14%) were performed on one election with sub-elections, and 11 studies (13%) were performed on many elections.

TABLE III. Number of Elections Characterized by Studies

Elections Studied	Studies
One Election	60 (72%)
One Election with sub-elections	12 (14%)
Many elections	11 (13%)

These data uncover another weakness in the current research: most studies are applied only once, in a very specific context. Thus, there is very little evidence of successful replicability of studies in other electoral contexts. In addition, the combination of this weakness with the lack of comparison to results of previous research, presented in Section IV.C, turns the results of many studies questionable.

In terms of the coverage of the elections, 68% of the studies related to national elections, 17% related to state elections, 8% related to municipal elections, 6% related to multiple elections and Tilton's pioneer study related to a university election. Also, in terms of the election role, the studies mainly considered presidential elections (32% regarding presidential elections and 10% regarding presidential primaries), and Parliamentary elections (35%). Senate, mayoral, and other types of elections corresponded to 16% of studied elections, and 7% of studies were related to multiple roles. Moreover, in 61% of the studies the vote was direct to a candidate, in 35% the vote was to a party, and 3 studies presented mixed types of vote. Finally, due to the prevalence of U.S. elections, the majority of studies focused on two-candidate races (42%), followed by the study of 3–5 candidates (30%), 6–10 candidates (11%) and elections with more than 10 candidates (6%). One study tracked and analyzed only one candidate and 11% of studies considered multiple scenarios. Fig. 3 summarizes the general characteristics of the studied elections.

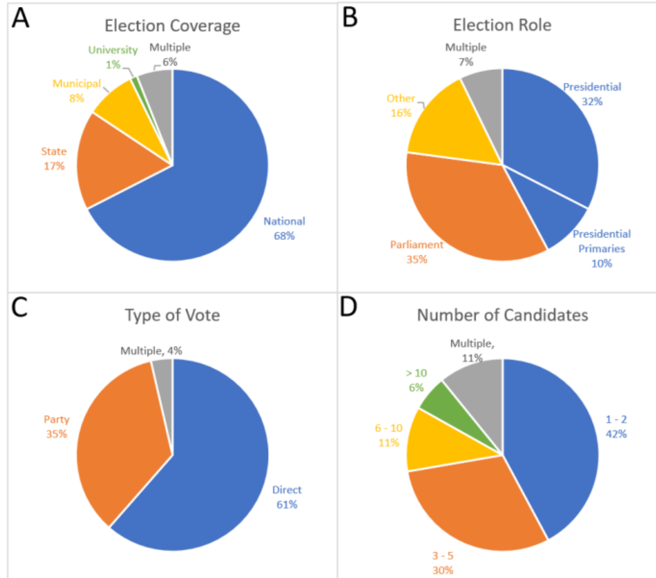

Fig. 3. Characteristics of studied elections: (a) coverage, (b) role, (c) type of vote, and (d) number of candidates.

The bulk of the studied elections were U.S. elections, with 30 studies related to the U.S., followed by India (11 studies), Taiwan (6), Pakistan and Indonesia (5), as shown in Table IV. In total, research attention is more focused on U.S. elections, which corresponded to a third of studied elections. Moreover, in terms of the specific electoral context, we identified 50 individual studied elections in 26 countries, ranging from 2008 U.S. elections to 2019 Indonesian elections, held earlier in 2019, as shown in Fig. 4. From the graph, it is possible to note an increasing trend to study U.S. presidential elections, starting with 3 studies in 2008, 10 in 2012, and 13 studies regarding the 2016 elections. Also, it is worth highlighting that there were three analyses of Taiwanese elections both in 2014 and 2016, and analyses of elections in India every year between 2013 and 2017. The first studied elections in Africa were the 2019 Nigerian elections.

The presented data brings attention to the possible existence of another bias in the studies' results. The most studied election context, U.S. presidential elections, presents a specific scenario with specific characteristics, such as the indirect relation between vote share and elections results, the existence of only two main political parties (Republicans and Democrats), and the concept of safe states (those in which a particular party's victory is already expected) and swing states (those that can reasonably be won by either the Democratic or Republican presidential candidate). These characteristics make these elections very specific, and results on approaches designed for these elections may be hard to replicate in other scenarios. For instance, in most Latin American countries, the presidential elections are raced by many candidates, the vote is direct, the concepts of safe or swing states do not exist, there are many parties, and even a small party may sometimes elect the president. Finally, the small number of studies related to

Latin America (only 8%) and Africa (only two studies), suggest that very few claims can be generalized to these regions.

Summarizing the answers to this research question, we first identified that most studies (72%) were performed in the context of a unique election, which can impact the applicability of their results, due to a lack of generalization. In addition, we identified that most studies are related to elections at a national level (68%), for the presidential position (42%), in a direct vote (61%) to a candidate. Also, usually there are only two candidates (42%), or a maximum of five candidates (72%). These data are in line with the most studied scenario: U.S. presidential elections. It is important to highlight, as discussed, that the prevalence of U.S. presidential elections may bias results, due to the specific characteristics of these elections, and the small number of studies regarding elections in Africa and Latin America shows that few assumptions can be made about elections in these regions.

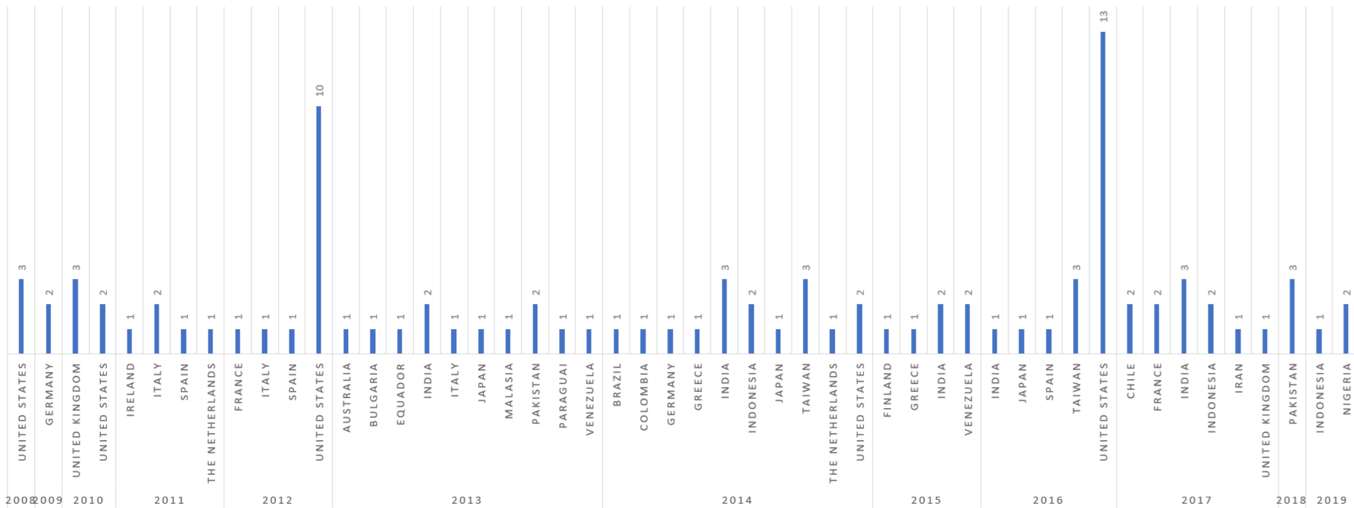

Fig. 4. Studied elections by year and country

TABLE IV. NUMBER OF STUDIES REGARDING ELECTIONS IN EACH COUNTRY

Studied Country	Studies
United States	30
India	11
Taiwan	6
Indonesia, Pakistan	5
Italy, United Kingdom	4
France, Germany, Japan, Spain, Venezuela	3
Chile, Greece, Nigeria, The Netherlands	2
Australia, Brazil, Bulgaria, Colombia, Equador, Finland, Iran, Ireland, Malasia, Paraguai	1

B. RQ2: What are the main approaches?

The analysis of extracted data identified the main characteristics of studies.

1) Social Networks used as input data

Despite claims to use SM for predicting elections, the vast majority of studies (73 – 88%) are narrowed to use only one social network as input. Twitter was the only SN used for most studies (62 studies, 75%), followed by Facebook (7 studies, 8%). Also, as presented in Table V, Twitter and Facebook together, combined or not with other input data (such as YouTube, Google Trends, blogs, candidates' or campaign pages), were used by six studies (7%); Twitter, combined with others but without Facebook, was used by two studies (2%), as well as Facebook, combined with others but without Twitter (2%). Finally, only four studies used neither Twitter nor Facebook: two used the Taiwanese online forum PTT Bulletin Board System, one used Flickr and one used Nairand, a social network targeted at Nigerians. Surprisingly, Instagram, one of the most popular social networks in 2020, was not used as input in any study.

TABLE V. SOCIAL NETWORKS USED AS INPUT DATA

Social Network used as data input	Studies	%
Only Twitter	62	75%
Only Facebook	7	8%
Twitter, Facebook and others	6	7%
Twitter and others	2	2%
Facebook and others	2	2%
Others (PTT, Flickr, Nairand)	4	5%

These data bring attention to two possible biases in the studies. First, there are many social networks, and it is hard to conclude that the study of only one social network (the scenario of 88%) is significant and generalizable to all social media scenarios. Second, the focus on Twitter alone cannot be justified methodologically. First, it is hard to find a discussion in papers that explains why Twitter was chosen. Second, Twitter is currently not even the most used social network. A recent report [36] estimated that, in July 2020, Twitter had 326 million active users, but Facebook had 2.6 billion active users and Instagram had 1.1 billion. Thus, future studies using Facebook and Instagram as social media, or a combination of two or all three SNs, are preferable.

2) Data collection

Regarding data collection, although some studies did not provide a detailed report on collection procedure, we next report and highlight the main findings.

On Twitter, data collection is usually performed through the official Twitter application programming interface (API) [41], either directly or using third-party services. User posts are collected by an open search on the Twitter platform, with previously specified filters defined according to three main criteria: (a) keywords, such as candidate or party names; (b) campaign hashtags used by candidates or parties; or (c) general hashtags related to elections. This data collection method is controversial, as it depends on keyword choices made by researchers, as discussed in Section VI. The number of collected tweets varied from 259 [PS39] to 400 million [PS34], with an average of 12.9 million tweets and a median of 250 thousand tweets.

In terms of the duration of data collection, it varies from 2 days to 7 months (220 days) of collection, with an average of 52 days and a median of 36 days, showing that studies usually collect data over a period of 1-2 months. Moreover, data collection was performed very close to election day. In 82% of studies with this information, data collection occurred until one or two days before elections, or even continued after election day. In studies using Twitter as source, usually the data collected consisted of tweet text with metadata, such as information about user, date, and time, and in some cases, the number of retweets.

On Facebook, data collection was mainly performed on data collected directly from official candidates or parties' pages, and, in one study, from supporters' groups. Studies that reported the number of pages monitored showed the following numbers: three studies monitored 3 pages, and four studies monitored 104, 222, 1126, and 1300 pages each. Collection period ranged from 1 to 206 days (average of 74 days and median of 75 days), and apart from one study that collected data until 3 days before elections, all others with this information (11) collected data until 1 day before or after election day. Unlike Twitter data gathering, data gathered on Facebook is not focused on the text of messages, but on metrics of public interaction. Collected data are usually the number of interactions on candidates' or parties' posts, such as the number of likes, comments, and shares, as well as the number of followers on official pages.

Other data collection methods were also identified in a few studies, such as a search for articles with candidates' names on PTT forum, search by keyword on Flickr, web scraping of posts on Nairaland, metrics from Google Trends with candidates' names, page views on Wikipedia entries related to candidates, and pages' traffic data from Alexa service.

3) Prediction approaches

After analyzing extracted data, identified approaches were grouped in five supermodel groups: (i) volume or sentiment; (ii) regression or time series; (iii) profile or posts interactions; (iv) topic analysis; and (v) other unique approaches. Table VI shows the number of studies classified according to each approach. The sum exceeds 100% because many studies use mixed approaches, and the table also shows the number of studies that use solely volume or sentiment approaches not combined with anything else.

Volume or Sentiment: More than three-quarters of studies (64 studies, 77%) are based on detection of volume and/or sentiment of text on SM. This is the main approach used by studies that included Twitter (61 out of 70 studies), and only three studies not based on Twitter used this approach, one based on Flickr data, other based on PTT data, and the third based on Nairand data. Moreover, this model is used as solely predicting technique by almost half of all studies (41 studies, 49%).

TABLE VI. MAIN APPROACHES USED

Prediction Model	**Studies**
Volume or Sentiment	64 (77%)
Volume or Sentiment solely	41 (49%)
Regression or Time Series	18 (22%)
Profile or Posts Interactions	14 (17%)
Topic Analysis	6 (07%)
Other approaches	6 (07%)

Studies using volume or sentiment modeling followed the proposal advanced by seminal studies such as those of Tumasjan [5] and O'Connor [6]. The former claimed that "the mere number of tweets mentioning a political party can be considered a plausible reflection of the vote share and its predictive power even comes close to traditional election polls" whereas the latter found that "a relatively simple sentiment detector based on Twitter data replicates consumer confidence and presidential job approval polls," as discussed in Section II. Thus, studies based on these approaches follow this process: (i) Twitter data collection by pre-selected keywords; (ii) data cleaning, by removing tweets not dealing with vote opinions and removing duplicates or retweets; (iii) sentiment analysis; (iv) prediction based on sentiment analysis counting combined with a simple linear formula; and (v) performance evaluation.

New studies using sentiment analysis usually have one or more of these objectives: (a) to replicate Tumasjan or O'Connor's studies in other (or even in the same) electoral contexts; (b) to improve prediction results by applying different sentiment analysis techniques; or (c) to improve results by tuning the prediction formula or data cleaning. The most used sentiment analysis techniques are lexical analysis, in which scores are applied to words previously classified as positive or negative; and machine learning models, notably Naïve Bayes [42], and Support Vector Machines (SVM) [43]. To improve a prediction formula, studies propose minor adjustments, such as considering only positive mentions, the difference between positive and negative mentions, or considering only one mention by a given user. In terms of data cleaning, some studies try to filter out possible bots, and others try to filter tweets by geolocation.

Most of the studies using this approach (42 studies out of 64, 66%) performed what we call one-shot predictions, *i.e.* just one prediction before elections; only a small group of studies (22 studies, 34%) was able to perform three or more predictions before elections, demonstrating the ability to be used actively during the campaigns.

Advantages include the facts that it is a simple counting approach, has a low cost, is easily implemented, and generates fast results. Also, many authors have claimed success or promising results using this approach in varied scenarios.

Many authors have also claimed negative results, as discussed in the next section, and some drawbacks persist since the initial studies. We highlight two main challenges. First, the majority of studies focused on the improvement of sentiment analysis, and not actually on the improvement of prediction. Lexicon-based analysis based on the presence of positive/negative words is very common, although previous reviews showing its limitations in this context [29][30]. Conversely, more sophisticated technics based on the advances of artificial neural networks (ANN), including recurrent neural networks (RNN) or deep learning, are almost never used and can be found in only three studies [PS02], [PS23], [PS61].

The second main challenge is that the nature of this model leads to many biases, including the following: (i) Twitter cannot be generalized as a good sample of all SM; (ii) collected data do not represent even a good sample of all tweets, due to platform constraints; (iii) it is too dependent on arbitrary decisions, such as search keywords and the selection of a period for data collection; and (iv) results are easily affected by volume manipulation from automated software, spammers, paid propaganda or even natural differences between users' online behavior.

Regression or Time Series: Regression and time series studies were grouped together because to most time series models are, or share characteristics with, regression models. This was the second-most identified approach, present in 18 of the 83 studies (22%).

The main characteristic present in most of these studies (13 out of 18 studies, 72%) is the use of traditional polls as additional input data, usually used as ground truth for training predictive models. Moreover, many of these studies (8 studies, 44%) were not only based on Twitter data, but also data from Facebook, Google Trends, Wikipedia and candidates' home pages. Thus, new variables, such as Facebook likes and comments on official profiles' posts, number of page views, and metrics of Google Trends were added to Twitter volume and sentiment, as well as retweets, to generate new sets of metrics. Then, these metrics were combined with offline poll results to train regression or time series models, capable of making predictions based on new instances of input data.

Use of traditional linear regression models, such as least square, ridge, and lasso, were found in eight of the studies in this category (44%). Moving average models, such as simple moving average (SMA), auto-regressive moving average (ARMA), or auto-regressive integrated moving average (ARIMA) models, were used by six studies (33%), usually combined with other approaches to generate continuous smoothed predictions. Other models were also found to have been used in a few studies. Kalman Filter and Gaussian process were used in two studies each, one study applied competitive vector auto regression, another created a regression based on what they called momentum, and one study tested a set of three high-dimensional methods: random forest, SVM, and elastic net. These metadata are summarized in Table VII.

Finally, in opposition to what was observed in the volume/sentiment approach, most of the studies using the regression/time series approach (72%) performed continuous forecasts over time, suggesting that this model may be most suitable for active use during the campaigns.

TABLE VII. TECHNIQUES USED BY REGRESSION OR TIME SERIES MODELS

Regression / Time Series models	**Studies**
Linear regression (least squares, rigde, lasso)	8(44%)
SMA / ARMA / ARIMA	6(33%)
Kalman Filter	2(11%)
Gaussian Process	2(11%)
Other	3(17%)

As advantages, these studies used machine learning and statistical methods for prediction. These methods are robust, well-grounded, and well tested in many other domains [44], [45]. Also, by using traditional polls as ground truth for training, results are less affected by volume manipulation. Also, the use of more SNs can reduce the inherent bias involved in using only Twitter as a data source and focusing on official profiles reduces the bias regarding keyword selection. Finally, this model seems to be more suitable for continuous predictions during the campaigns.

In terms of challenges, some biases can also be identified, such as the arbitrary selection of data sources, collected data, and the period of collection. For example, a different window size on the moving averaging techniques can totally change the results. Also, models chosen for regression and time series are limited for this context: linear regression may be not suitable, due to a possible nonlinear relationship between SM variables; and the ARIMA model is univariate, and therefore does not allow the combination of multiple variables. In consequence, many studies analyzed each metric individually and chose the one with the best results, an experimental procedure to be considered with caution.

Profile or Posts Interactions: The number of interactions on posts or on the official profile of candidates or parties was also considered in some studies (14 studies – 17%). Three types of studies used this approach: (i) studies considering Facebook likes on posts made by official profiles as approval rate or vote intention, in a similar way to how volume/sentiment approaches used mentions on Twitter; (ii) studies using a similar approach considering likes and dislikes on the Taiwanese PTT Bulletin Board System; and (iii) studies using likes or retweets as additional metrics in volume or sentiment models. These studies basically considered new metrics for prediction, not imposing novelty on the prediction model.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TCSS.2021.3063660

Topic or Event Detection: Topic or event detection and analysis are also supportive methods for other already mentioned approaches. In the six studies using this approach, they were used mainly as support for or as a replacement of sentiment detection. By using Latent Dirichlet Allocation (LDA)[46], studies attempted to find the most important subjects being talked about in an election, the alignment of these topics with candidates, and then the volume and sentiment of public posts for or against the candidates. These studies may be considered as specializations of volume/sentiment approach, sharing their other characteristics.

Other Approaches: Unique studies include approaches based on prediction market, cluster detection, centrality score, statistical physics of complex networks, and analysis of groups of supporters, solely or in combination with previously described approaches.

4) Prediction procedure and evaluation metrics

In line with the data collection procedure, in which most studies collected data until election day or 1-2 days before elections, most studies (71 – 84%) tried to predict election results one day before elections, 10 studies (12%) made predictions between 3 and 10 days before the election, and four studies made predictions on dates ranging from 49 days to 8 months before elections, as shown in Table VIII. The sum is higher than the number of studies because some studies used mixed approaches.

TABLE VIII. NUMBER OF DAYS BETWEEN PREDICTION AND ELECTIONS DATES

Days between prediction and elections	Studies
1 day before	71 (84%)
3 - 10 days before	10 (12%)
> 10 days before (49 - 253 days)	4 (5%)

Despite the fact that most studies made predictions one day before elections and most studies performed one prediction total (54 studies, 65%), some studies (29 – 35%) pursued a more detailed approach and made daily predictions or many predictions over time. Thus, those approaches, if successful, may be able to predict not only the final election results, but also to perform nowcasting, which refers to making prediction on a daily basis, using SM to capture variations in vote intentions throughout the campaign. Thus, this kind of approach sounds more useful because it can be used as a complement to traditional polls during the campaign period.

Despite approaches having been designed to predict election results one or several days before elections, only eight of the studies (10%) claimed to have made predictions or publicized any kind of result before official election results. Authors of these papers claimed that they had publicized results from between 1 day up to 8 months before elections. These data, regarding 90% of studies being performed after elections, can leverage the existence of bias on results' validity due to, as aforementioned and also expressed in [22] and [30], many results being contingent on arbitrary choices of the authors and being unintentionally biased if choices were made after knowing target results or by just selecting the model with best results.

The most used prediction metric was vote share (64 studies, 77%), followed by just noting the election's winner (17 studies, 20%) and six studies (7%) used other metrics, such as the order of candidates. In terms of error metrics, a surprising result was found. Many studies (38 studies, 46%) used no error metric at all, analyzing results as just "very close to" or "far from" final elections result in a dubious assessment. Moreover, the most used metric (33 studies, 40%) was mean absolute error (MAE), already noted by election forecasting literature [47] as a suitable error metric for this scenario. Also, nine studies (11%) used root mean squared error (RMSE) and the same number of studies used other metrics, such as absolute error (AE), absolute percentage error (APE), or statistical correlation with results, as shown in Table IX. Results exceed 100% because some studies use more than one prediction and error metrics.

Finally, it was also checked which studies performed statistical tests on results, to verify whether they were statistically significant. Only seven studies (8%) were found to have performed any kind of statistical tests on results. Statistical tests performed included the Wilcoxon signed-rank test, Wilcoxon–Mann–Whitney test, Welch's t-test, and paired t-test. It is worth noting that without statistical tests in most studies it is hard to know whether the results they claimed were either statistically significant or were obtained merely by chance.

TABLE IX. ERROR METRICS USED

Error Metric	Studies
Simple Comparison	38 (46%)
Mean Absolute Error (MAE)	33 (40%)
Root Mean Squared Error (RMSE)	9 (11%)
Other (Statistical correlation, AE, APE and others)	9 (11%)

Results exceed 100% because some studies use more than one error metric

C. RQ3. What are the main characteristics of successful studies?

In this section, we try to identify correlations between studies' characteristics and successfulness. An exploratory, descriptive analysis of each variable was performed for this purpose.

For the analysis, extracted data were modeled as follows. We considered as successful the studies directly claiming good results but excluded those with MAE higher than 6% on vote share and other questionable results, such as success on predicting vote share of one candidate but failure on others. Our definition of the threshold of 6% was based on a recent study [47] that has analyzed more than 30,000 polls between 1942 and 2017 and found a historical MAE of 2%, with almost no variation over time. Thus, we considered as not successful the studies with MAE higher than three times this historical value. Other mixed results, such as having achieved success in one round but failure in another, success at a national level but failure at a state level, or success in one country but failure in another, were categorized as "not clear" results for the approach.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TCSS.2021.3063660

Some papers reported results in more than one election. For these, if the result was the same in all contexts, we considered the average contextual numbers, *e.g.*, number of candidates and volume of data gathered. Also, some extracted data became new binarized variables, such as models and SNs used, because the same study may have used more than one model, as well as SN. Due to this data handling, some numbers presented in this section are different from those presented in previous sections. Next, we present the analysis.

Less than two-thirds of studies (52 studies - 63%) were considered successful studies, 28% (23 studies) were considered unsuccessful, and 10% (8 studies) of studies were categorized as having no clear results. Given that this type of research encourages the reporting of positive results, the low success rate of 63% is alarming and puts in doubt the purported feasibility of predicting elections based on SM data. Also, if we consider the already discussed methodological limitations of studies as the lack of replication in more than one context and the lack of statistical analysis, it is plausible to consider that success can be obtained merely by chance, as directly argued in some studies. Furthermore, it is essential to stress that no assessment of technical adequacy of the models to the context has been analyzed here, because this task was out of the scope of this review.

By trying to find evidence that correlates studies' characteristics and success, we highlight some results that reflected success in a baseline above the 64% of general success.

Aligned with our argument that methods should be tested in different electoral contexts, four out of five studies applied in three electoral contexts obtained success. Regarding election role, primaries (75%), parliamentary (69%) and presidential (63%) elections obtained better results than elections for mayor, senate, and governor elections (aggregated 47% success over 19 studies). There is a small difference regarding vote on party (69%) or direct on candidate (63%), but there is a notable difference regarding electoral year: in 35 studies regarding elections occurring between 2012 and 2015, there was 77% success, in contrast to 47% of 15 studies related to years between 2008 and 2011, and 55% of the 33 studies on the years between 2016 and 2019. These data show that, in opposition to expectations, the success rate of studies does not increase over time.

Due to the fact that many countries received attention in just a small number of studies, we grouped countries by continent and by economic development, according to United Nations classification [48]. As result, we found that studies on Asia (73% of success) and Latin America (71%) performed better than studies on Europe (63%) and Anglo-America (54%), despite the prevalence of studies being performed on the U.S. Moreover, studies on developing economies achieved more success (74%) than on developed economies (57%), challenging the conclusions presented by Koli [34], who argued that predictions yield better results in developed countries.

In terms of the approach used, the use of a volume or sentiment was not a good approach: only 55% of the 64 studies that used this approach obtained success, in contrast to 89% of the 19 studies that did not use volume or sentiment. Reinforcing this finding, 72% of the 18 studies that used regression or time series approaches, 64% (of 14 studies) that used profile or posts interactions, 83% (of 6 studies) that used topic analysis, and 83% (of 6 studies) that used other specific approaches obtained success. These data allow us to argue that, despite being the most used approach, the volume and sentiment approach is probably not the best way to predict elections based on SM, and more research should be done on other approaches, in special regression and time series, and topic analysis approaches.

In line with the previous conclusion, Twitter is not the best platform for data collection. While 60% (of 70 studies) based on this SN were successful, 77% (of 13) not using Twitter achieved success. Moreover, better results were achieved with other platforms: 80% of studies based on Facebook were successful (against 59% of studies not using Facebook data), as well as 85% of studies using other data sources. Additionally, using polls to train the models also appears to be a promising practice: 76% of the 17 studies using polls as a data source were successful, compared to 59% success among studies that did not use polls.

Regarding the number of days of collection, there were no significant differences found. Studies that collected data for less than 31 days, between 31 and 90 days, and for more than 90 days achieved success rates of 61%, 64% and 64% respectively. Finally, regarding the volume of microdata collected (e.g., number of tweets or Facebook posts), better results were obtained when a high volume of data was collected. From 47% success of the 43 studies that collected less than 500,000 data points to 73% of success of the 26 studies that collected more than 500,000 data points.

A summary of this analysis is presented on Table X, including the characteristic, total number of studies with the characteristic, and the success rate.

VI. Discussion of Main Strengths, Challenges, and Future Directions

In this section, we aim at answering the last research question "RQ4. What are the main strengths and challenges of predicting elections with social media?," by summarizing and discussing the results presented in the previous sections. Thus, we present possible future directions for studies in this area.

A. Main strengths of predicting elections with social media data

As main strengths of analyzed studies, we can highlight:

Use of new large amount of available data: There is a large amount of data available on SNs, including data about what people are saying about politicians or political parties, what politicians and parties are talking about, and repercussion and reach of conversations. This data availability is unprecedented in human history and has changed the concept of media influencers. The change is from an era when the influence was mainly enjoyed by "big players" present on traditional media, mainly TV, to an era when ordinary people

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TCSS.2021.3063660

in small cities, with low or no budget, are able to exert significant influence.

TABLE X. Average Success Rate

Characteristic	Total of Studies	Success Rate
One Electoral Context	72	63%
Two Electoral Contexts	6	50%
Three Electoral Contexts	5	80%
Election Role: Presidential	27	63%
Election Role: Presidential Primaries	8	75%
Election Role: Parliament	29	69%
Election Role: Other	13	46%
Election Role: Multiple	6	50%
Type of Vote: Direct	51	63%
Type of Vote: Party	29	69%
Type of Vote: Multiple	3	0%
Number of candidates: 1 - 2	35	66%
Number of candidates: 3 - 5	25	64%
Number of candidates: 6 - 10	9	67%
Number of candidates: > 10	5	20%
Number of candidates: Multiple	9	67%
Election Year: 2008-2011	15	47%
Election Year: 2012-2015	35	77%
Election Year: 2016-2019	33	55%
Continent: Asia	26	73%
Continent: Anglo-America	24	54%
Continent: Europe	19	63%
Continent: Latin America	7	71%
Continent: Africa	2	50%
Continent: Oceania	1	100%
Continent: Multiple	4	25%
Economic Status: Developed	47	57%
Economic Status: Developing	34	74%
Economic Status: Multiple	2	0
Overall Success Rate	83	63%

Characteristic	Total of Studies	Success Rate
Model: Volume or Sentiment	64	55%
Model: Not Volume or Sentiment	19	89%
Model: Regression or Time Series	18	72%
Model: Not Regression or Time Series	65	60%
Model: Profile or Posts Interactions	14	64%
Model: Not Profile or Posts Interac.	69	62%
Model: Topic Analysis	6	83%
Model: Not Topic Analysis	77	61%
Model: Other	6	83%
Model: Not Other	77	61%
Social Network: Twitter	70	60%
Social Network: Not Twitter	13	77%
Social Network: Facebook	15	80%
Social Network: Not Facebook	68	59%
Social Network: Other Networks	13	85%
Social Network: Not Other Networks	70	59%
Other data input: Polls	17	76%
Other data input: Not Polls	66	59%
Days of collection: 1 - 30 days	33	61%
Days of collection: 31 - 90 days	28	64%
Days of collection: > 90 days	14	64%
Days of collection: Not specified	8	63%
Data volume: Less than 100,000	22	50%
Data volume: From 100,000 to 499,999	21	43%
Data volume: >= 500,000	26	73%
Data volume: Not specified	14	93%
Overall Success Rate	83	63%

The number of studies may have small differences those presented in previous sections because of the handling of studies performed in more than one election.

Real time data availability, collection, and analysis: In addition to having such a large amount of available data, these data can be collected and processed in real time. This capability opens new opportunities in political campaigns, as these data may support quick adjustments in campaigns, policies, or speeches, *e.g.*, in real time during a debate.

Low cost: Due to automated data collection and analysis, these approaches can be considered as low cost, relative to traditional offline polls, when a coordinated operation of a high number of interviews is usually needed.

Advances of artificial intelligence: These approaches are strongly based on artificial intelligence. Fortunately, the last decade has seen substantial development in this area, including models and algorithms, and also available hardware for model training and prediction execution, such as GPUs, distributed systems, grid computing and cloud computing. Thus, computations that, a few years ago, took weeks to execute, may currently be executed in a few minutes.

B. Main challenges of predicting elections with social media data

As main challenges of the studies, we can highlight:

Lack of well-defined and replicable processes: It is hard to find, among the studies, the definition of detailed and replicable processes, explaining and justifying the options and choices, in a way that would yield replication in other scenarios by other researchers. Thus, as consequence, despite some efforts to replicate past results with data from another or even the same elections, the results achieved are usually quite different.

Lack of generalization: Combining the lack of replicable processes with the fact that most studies were applied to only one electoral context, there is little evidence that the proposed approaches are applicable in other electoral contexts or if they are generalizable. Thus, there is little evidence to determine whether positive results were obtained just by chance, by overfitting the model to that specific election, or because it was a feasible predictive model. Moreover, due to the focus of the majority of studies on U.S. elections and the specific characteristics of this electoral context, it is hard to envision the results of application in other contexts. For example, in [PS06], authors applied the same approach in U.S. and India, and obtained success in the former but failed in the latter.

Lack of prediction capabilities during the campaign: Almost all studies were performed after election results was made public, and most studies were designed to perform only "one-shot" predictions, *i.e.*, one prediction before elections, usually the day before. This design limits the applicability of approaches during campaign rallies, and there is little evidence that they are reliable for use during future campaigns. In fact, most studies can be considered as posterior analyses of how the behavior on SN correlated to election results with descriptive goal, instead of how to perform predictions during electoral rallies.

Social networks do not represent a good population sample: Social networks cannot be considered a good sample of the population and should not be used as the only input data capable of generating generalizable results. For example, a recent report [36] shows that only 51% of the world population use SN, the majority of whom are young people and men. Additionally, another report published in 2019 points out that Twitter users in the U.S. are younger, likelier to be identified as Democrats, more highly educated, and have higher incomes than U.S. adults overall [49]. Those data do not reflect world or U.S. demographics.

Twitter is most used but does not represent a good sample of social network: Twitter is the SN used in most of the studies (84%), and in many of them (75%), it was the only SN used as input. However, Twitter is not a good sample, even considering only SM users, due to its having very few active users (326 million), relative to other SN, such as Facebook (2.6 billion) and Instagram (1,1 billion), according to a 2020 report [36]. Despite these data, it is hard to find a discussion about why studies focused on Twitter. After analyzing the API of these SNs [41], [50], we hypothesized that Twitter was chosen because it is easier for researchers to collect data on this platform. For example, starting on August 2018, the approval process to gather data from Facebook and Instagram consisted of developing and deploying a fully functional system, creating and publishing a privacy policy and terms of use, recording a video showing all the functionalities related to Facebook and Instagram data collection, creating test accounts allowing Facebook employees to test the system and, in many cases, sending formal documentation of an institution responsible for the system. By contrast, in August 2019, the Twitter approval process only involved completing a form with information about the system.

Collected data on Twitter do not represent a good sample of Twitter data: Twitter API may be used in two ways: streaming or query. By trying to gather large amount of data, such as those used by Twitter-based approaches, developers can be limited in two ways: it returns a random sample of recent tweets published in the last seven days; or the user is limited to 180 calls (that returns a maximum of 100 results by call) for a window of 15 minutes, which is usually not sufficient to gather all tweets related to candidates.

Arbitrary data collection choices: In most studies, many data collection choices were arbitrary, such as the data collection period, which usually varied from 3 days to 3 months before elections, and the keywords used for open search on volume/sentiment approaches. This created many problems, such as those presented by [30], in which the performance was too unstable because it depended strongly on such parameterizations, and unintentional data dredging could occur, due to post hoc analysis. Also, it reinforces the argument presented by Jungherr [22] who, after replicating the seminal study of Tumasjan [5], argued that "the results are contingent on arbitrary choices of the authors," and indicated that simply including one more party or day of collection would greatly change the results.

High susceptibility to volume manipulation: Data volume manipulation on SN may be imposed in many ways, such as the use of automated software, known as BOTs [51], [52], spammers, paid propaganda, astroturfing, or even natural differences between users' behavior [53].

Difficulties in crossing data from multiple networks: The approach based on open search used in Twitter-based studies is hard, if not impossible, to implement on other social networks, due to limitations of the API. For example, Facebook and Instagram do not allow open search of general keywords. Also, even in studies considering high level metrics on regression or time series models, the models used are not suitable to perform data analysis in an aggregate way. Thus, in these studies each metric was analyzed and used for prediction in an independent way, not allowing for the crossing of data from multiple networks and limiting the effectiveness of results.

Lack of use of state-of-the-art machine learning: In studies based on volume or sentiment, the focus is more on the improvement of sentiment analysis, rather than on the improvement of the prediction model. Nevertheless, most studies relied on simple lexicon-based methods or on well-established methods, such as Naïve Bayes and Support Vector Machines (SVM). However, as already presented, these studies achieved little success. In addition, even in studies based on regression and time series, only simple and traditional methods were applied, with prevalence of linear regression based on least squares, ridge, or lasso algorithms, and SMA/ARMA/ARIMA models for time series. Linear regressions are meant to describe linear relationships between variables, which cannot be assumed in this context. Also, ARIMA is a univariate model, and hence cannot exploit the leading indicators, nor combine multiple features as aforementioned.

There are recent advances on machine learning models capable of dealing with these limitations, such as improvements on artificial neural networks, including recurrent neural networks or deep learning, but they are almost not yet considered in current studies.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TCSS.2021.3063660

Technical modeling weaknesses: Despite having been recognized by some authors that the electoral prediction may be considered a time series forecasting problem with very short series, the authors have yet to bring to the fore data preprocessing techniques and AI time series modeling for the SM environment.

That involves the precise problem characterization, the underlying mathematics of its dynamics and the approximations needed in the data analyses and preprocessing. This review was not able to reach papers in these topics, and concepts such as homoscedasticity, techniques such as RFM-Analysis and performance metrics such as Theil's U are yet to be considered in this interdisciplinary field.

Additionally, it is well known that the results of using AI techniques and models can be very affected by chosen parameters. However, very few studies take this in account, and the vast majority do not even mention anything about what parameters were used. From the studies that mention this aspect, 3 reported the use of default parameters of used tools, namely Weka [54] and Scikit-learn [55] and 4 make clear that they chose parameters by "trial-and-error". Exceptions are [PS23] and [PS42], that presented discussions in this regard. This scenario presents a meaningful weakness in the area since failure may be related to the parameters' choice instead of the model itself.

Performance evaluation and scientific rigor: Additionally, the quality assessment and analysis of studies presented important drawbacks that can affect the results' reliability: lack of statistical analysis of results; lack of meaningful comparison of results with related works; and lack of discussion regarding biases and threats to validity present in studies. The lack of these analyses and comparisons, when added to other such challenges as lack of replicable processes and generalization, casts doubt on the actual prediction capabilities of approaches based on SM.

The challenges presented above may be grouped in four categories, (a) process, (b) sampling, (c) modeling, and (d) performance evaluation and scientific rigor, as summarized and presented in Table XI.

TABLE XI. SUMMARY OF MAIN CHALLENGES

Category	Challenge
Process	Lack of well-defined replicable processes
	Lack of generalization
	Lack of prediction capabilities during the campaign
Sampling	Social networks do not represent a good population sample
	Twitter is the most used but is not a good sample of social
	Collected data on Twitter is not a good sample of Twitter data
	Arbitrary data collection choices
Modeling	High susceptibility to volume manipulation
	Difficulties in crossing data from multiple networks
	Lack of use of state-of-the-art machine learning
	Technical modeling weaknesses
Performance evaluation and Scientific rigor	Lack of statistical analysis of results
	Lack of meaningful comparison of results with related works
	Lack of discussion regarding bias and threats to validity

C. Future directions

The results indicate that the research in this area is still in its infancy. Next, a discussion about its future is presented.

Future Directions in Process Definitions

As the most important direction for the future, we consider that studies should cease to be merely ad hoc initiatives and aim to become generalizable and repeatable processes. Thus, it may be possible to apply new approaches and models in many different electoral contexts, such as different countries and years, by proposing and testing improvements and comparing results. For this, processes defined for data mining and knowledge discovery may be used as a basis, such as CRISP-DM [56], SEMMA [57] or DMLC [58]. For example, CRISP-DM presents six phases: (i) business understanding, (ii) data understanding, (iii) data preparation, (iv) modeling, (v) evaluation, and (vi) deployment. Based on these phases, new approaches may benefit from detailing steps, inputs and outputs, models, and algorithms to be used in each phase, to become repeatable and generalizable.

Moreover, the process should also be adjusted to allow approaches to be used during campaign rallies, to increase its usefulness by opening new opportunities that support quick adjustment on campaigns, policies, or speeches in a continuous way.

Future Directions in Model Definitions and Sampling

We agree with the authors of [59], who stated that "researchers should refrain from automatically generalizing the results of single-platform studies to social media as a whole," and results show that studies covering multiple social networks are necessary to better frame the prediction scenario. Also, the research must have some characteristics. First, by using many SNs as input, studies should consider the different behavior of politicians and users on each platform. For example, one politician may have higher engagement on Twitter, but others may perform better on Instagram. In an extreme case, one candidate may perform better on one SN at the beginning of a campaign, but the behavior may change later. Second, data collection should be systematic and uniform in all involved SNs, to allow the combination of different SN data as input data, and to avoid the common bias of arbitrary choices made by researchers. Third, new models should be resistant to volume manipulation, such as that threatened by spam, paid propaganda, bots, or even different behavior of the electorate on the Internet.

As a possible direction, the use of state-of-the-art ML algorithms, for instance, based on ANNs may be a recommended approach, due to their characteristics: (i) ANNs can learn nonlinear mappings capturing complex relations among independent (input) and dependent (output) variables; (ii) ANNs do not need explicit assumption for the model between the inputs and outputs; (iii) ANNs can generalize well; and (iv) ANNs do not require assumptions on the distribution of input data, unlike most statistical techniques. In particular, the multilayer perceptron (MLP) is likely to be useful in this research for having extra features such as being the most validated ANN, easy to use and a universal function approximator [60]. Also, to avoid volume manipulation, the training of ML algorithms on traditional polls is already presenting promising results.

Also, the precise problem characterization, the underlying mathematics the problem dynamics and the approximations needed in the data analyses and preprocessing already used in the fields of time series forecasting must be addressed to leverage the quality of models. Finally, the proper addressing of precise parameter selection and tuning for models may also unlock a new level of reliability and robustness to the results.

Future Directions in Evaluation

To allow a better evaluation of the studies' results, future works may focus on establishing a common framework of evaluation and common baselines. As was well discussed by [61], success must be measured statistically, not merely through description or mean average error, and must be relative to clear benchmarks, which can be previous election results, existing polls, or default assumptions, such as incumbency success. Thus, the application of statistical tests, such as Wilcoxon signed-rank test, Wilcoxon–Mann–Whitney test, Welch's t-test, or paired t-test, just to cite a few, should be addressed.

Finally, studies' reports should clearly discuss bias and threats to validity, together with the results.

VII. Comparison with Previous Works and Limitations of This Study

In this section, we compare the methodology and results of this work and previous similar studies. Then, we discuss the threats to validity and the limitations.

A. Comparison with previous works

The search process found seven studies aimed at reviewing the literature regarding predicting based on social media data. Two of them [31] [32] just presented, in general lines, the main approaches for predicting using Twitter. The other five presented more detailed studies: Kalampokis *et al.* [29] and Gayo-Avello [30] in 2013, Kwak and Cho [33] in 2018, Koli *et al.* [34] in 2019, and Bilal *et al.* [35] in 2019. These five studies were already presented on Section II and some of their results were discussed throughout the text of this study. In this section, we discuss the main similarities and differences on methodology and conclusions.

An important difference from previous studies concerns methodology. Only the study [29] followed a systematic approach. Its search was performed on Google Scholar, using the keywords "predict OR forecast AND social media," and inclusion/exclusion criteria aimed at selecting papers focused on predicting real world outcomes. As a result, the study analyzed 52 papers performing predictions on Twitter in many domains, including 11 focused on electoral context. Despite being well-organized, the other four studies did not follow a systematic approach, and a rigorous method for paper selection, filter, data extraction and analysis is not clear. Also, only one study [33] analyzed a high number of studies (69), whereas all others analyzed less than 13 papers regarding election predictions. Also, two of them [30] [34] only focused on Twitter as SM, narrowing the results to this SN.

Also concerning methodology, this study covered a broader set of data. For instance, none of these previous studies performed a quality assessment or analyzed and summarized the electoral contexts. Also, the duration of data collection was only analyzed by [30], and data summarization and analysis of the volume of collected data was not found. Moreover, no analysis focused on discovering correlations between studies' characteristics and successfulness, as performed in this study, was found.

Regarding models, all of them recognized volume and sentiment models as the most used, and one recent study [34] also identified regression and time series models. Moreover, all of them challenged the prediction power of this volume/sentiment approach. In addition to these previous results, our study was also capable to identify two new supporting approaches, namely, profile or posts interactions and topic or event detection. Also, we identified new approaches, such as cluster detection and statistical physics of complex networks. Moreover, we presented a direct comparison of success between volume/sentiment approaches and the other existing ones, which had not been presented in previous studies.

In terms of similarities, some studies had already pointed out some challenges, such as those associated with the lack of replicable processes and generalization, the non-representative nature of Twitter as SM and collected data not being a good representative sample of Twitter data. Moreover, the bias related to high susceptibility to volume manipulation is a common conclusion. These challenges were also pointed out in our study, suggesting that, in some way, past issues are not properly addressed in recent studies.

Considering future directions, the future of regression approaches indicated by Gayo-Avello [30] started to be implemented and was found in our analysis. As an update, we add that, despite good initial results, the enhancement of ML techniques used on regression studies, as well as the training with polls, may be a promising future direction. Moreover, the ability to fuse data from multiple networks was neither addressed in past studies, nor was there discussion of the definition of processes based on existing methodologies such as CRISP-DM, and the need to use statistical tests, such as Wilcoxon tests.

Finally, the main results of [34] were challenged in our study. After analyzing 13 studies, Koli *et al.* claimed that:

> *"1. One can predict the election outcome with Twitter Data only in developed countries like Germany, France, USA and Italy. Because in these nations, the literacy rate is above 99% and more than 80% population access Internet.*
> *2. One cannot predict the election results with Twitter Data in developing nations like India, Pakistan, Sri Lanka, etc. Because in these nations average literacy rate is below 70% and only small portion of the population uses the Internet (especially Twitter less than 10%)."*

However, after analyzing 83 studies, we found that 74% of studies on developing countries achieved success, but only 57% of studies on developed countries succeeded, and this is the opposite of Koli's result.

B. Threats to Validity and Limitations

Despite the rigor with which this study was conducted, it may have been affected by threats to validity, particularly with regard to finding all the relevant studies, assessing their quality, and extracting data.

Given the increasing number of studies in the area of predicting elections with SM data, there is no guarantee that all the relevant studies were identified. Even though we applied a rigorous search, described in Appendix I, some papers may have escaped inclusion. Although four digital libraries were selected as sources, which we assumed would include all high-quality relevant studies, they are not exhaustive. To minimize this issue, we applied a mixed approach to find relevant studies that combined an automatic search in search engines, and a snowballing search, that is, searching for relevant studies in the references of previously selected studies. Moreover, the findings are based on papers published in English only, as all non-English studies were discarded, and purely on academic peer-reviewed publications, limiting the scope to academic studies and not considering the knowledge reported in other sources, such as technical websites, blogs, etc.

Quality assessment and data extraction tasks were individually performed by each of the two reviewers and disagreements were discussed in a consensus meeting. Although this procedure increased our confidence in the reliability of this study, we nonetheless found that quality assessment and data extraction may have been compromised by the way most of the studies were reported. The report organization of some studies made it difficult to locate the required information in the extraction process. Furthermore, many papers did not present sufficient information, and, in many cases, information had to be inferred from the text. Therefore, despite the effort to reach a consensus during data extraction and quality assessment, there may have been some inaccuracies in the inferred data. Finally, some studies were performed in more than one electoral context. We are aware that if we had decided to consider each applied context of all studies as a different study, some success percentages would be different.

VIII. Concluding Remarks

This study collected more than 500 articles, 90 of which were focused on predicting elections based on SM data, investigating, and summarizing how this new research field has evolved since 2008. Among these studies, 83 are primary studies aiming at predicting elections and seven are surveys or reviews of past studies.

The results show that the number of publications in this area is increasing and research is spread across 28 countries from all continents. Nevertheless, there cannot yet be found any prominent researchers, research groups, or clusters performing sustainable research in the area. Also, there was no identification of a common well-known forum for publication on this subject, and results are spread across many forums.

Regarding electoral contexts, most studies were performed in the context of a unique election, which may impact the results' validity. Also, most were related to presidential elections at a national level with few candidates. Moreover, the most studied scenario was the U.S. presidential scenario, which can impact generalization due to its specificity.

Considering the main models used, we found that most studies used the approach of volume/sentiment analysis only on Twitter, in a variety of data collection approaches. We also found that regression and time series analysis is increasing, using multiple SNs, in addition to some supporting approaches, such as profile or post interactions and topic analysis.

By combining studies' characteristics and success we found that, despite being the most used approach, volume/sentiment does not present high success rates, which is consistent with the conclusions of previous surveys. Thus, approaches such as regression or based on profile/posts interactions may be better to investigate and improve; even totally new approaches, such as one based on statistical physics of complex networks, may be tested. Finally, studies based on Twitter achieved significantly lower success rates than studies based on other SNs, such as Facebook. Surprisingly, no studies based on Instagram were found.

Moreover, as main challenges, we identified issues in four areas. Regarding processes, we highlight the lack of well-defined, replicable and generalizable processes, and lack of prediction capabilities during the campaign. In sampling, issues are mainly related to the fact that SNs and Twitter data do not represent representative samples, and studies were performed with many arbitrary data collection choices. Regarding modeling, we found difficulties crossing data from multiple networks, the high susceptibility to volume manipulation, the lack of use of state-of-the-art ML techniques and technical modeling weaknesses. And considering performance evaluation and scientific rigor of studies, the lack of statistical analysis of results and of meaningful comparison with related works are also main issues.

Finally, the study presented the authors' point of view on the future directions of predicting elections using SM data in three axes: process definitions, model definitions and sampling, and study evaluation. As main directions, we highlight the need for repeatable processes based on well-known methodologies, for example CRISP-DM or SEMMA; the use of state-of-the-art methods for regression based on machine learning that can combine data from multiple SNs, such as ANN; and the use of statistical tests for results evaluation, such as Wilcoxon signed-rank test and others.

The results from this review contribute to the research field by providing the academic community, as well as practitioners, with a better understanding of the research landscape and by identifying some of the gaps in the area that open up opportunities for future research. In addition to future directions presented, this literature review may also be extended in certain ways: a search extension may be performed to expand the search strategy and number of sources, thereby performing a broader study; a temporal update can be implemented without making modifications to the protocol, to expand the timeframe and compare results over time; and finally, both approaches can be combined.

APPENDICES

I. DETAILED METHODOLOGY

The method chosen for this research was a systematic literature review, which has proven to be a replicable and effective manner to identify, evaluate, interpret and compare studies that are relevant to a particular question or area, and is widely used in some research areas, such as software engineering and health care [37]–[40]. Following the guidelines defined by [40], the method used in this research is defined below.

A. Research Questions

To define the research questions of this study, we returned to the main objectives:

To provide a thorough review and investigation of the state of both the art and the practice of predicting election outcomes based on SM data and to identify key research challenges and opportunities in this field.

Then, the following research questions were derived:

- RQ1: In which electoral contexts is the research being performed?

This question aims at identifying the electoral contexts being studied, such as the year and country in which the election took place, and the type of election. This question is intended to ascertain whether the studies are best suited or giving attention to any particular electoral context.

- RQ2: What are the main approaches?

The objective of this question is to identify the main approaches used, their main characteristics, how they are modeled and applied to predict elections, and which are the metrics used to assess their performance.

- RQ3: What are the main characteristics of successful studies?

The objective of this question is to identify the main characteristics of allegedly successful studies, in order to identify in which specific contexts, which approaches, and which factors yield effective results.

- RQ4. What are the main strengths and challenges of predicting elections with social media?

After studying the context, approaches and characteristics of successful studies, the answer to this question aims to summarize the main perceived strengths, weaknesses, challenges, and opportunities in this new research area to guide future research.

B. Research Team and Decision Procedure

A team of three researchers developed this study, two of whom, Kellyton Brito and Rogério Silva Filho, are undergoing their Ph.D. studies and composed the reviewer group, and Paulo Adeodato, who is a full-time lecturer and supervised all activities. Regarding the two students, one has research and practice experience in the domain of politicians' data and open government data and is also experienced on systematic literature reviews and data analysis; the other is experienced on data mining and machine learning. Regarding the supervisor, he is an expert in the research and practice of data mining and machine learning, working with predictions in many different domains for more than 25 years.

All team members were involved in defining the scope, research objective, research questions, and methodology, as well as discussing the findings. The review process was implemented by the reviewer group, under the direction of the supervisor.

Important activities in a systematic study may lead to conflicts that require decisions to be made regarding study selection, quality assessment, and data extraction. It is thus recommended that such activities be performed by at least two researchers. To address these situations, and to diminish threats to validity, for this study we defined the following decision and consensus procedure: During the review process, the decision procedure began with both researchers of the reviewer group individually performing in redundancy all activities related to the study selection, quality assessment, and data extraction. After individual evaluation, the results were integrated into an agreement/disagreement table and a decision meeting was held. During the decision meeting all the results that had at least one disagreement were discussed by the members until a final consensus was reached.

C. Search Process

The rigor of the search process is one of the distinctive characteristics of systematic reviews [38]. To implement an unbiased and strict search, two approaches were combined: (i) automated search on indexing systems, and (ii) snowballing search on the references of studies found on the automated search.

The automated search was performed in four indexing systems: ACM Digital Library, IEEEXplore Digital Library, ISI Web of Science, and Scopus. The search was performed on papers' metadata: title, abstract and keywords and aimed to find studies focused on predicting elections based on SM data. Then, after some initial refinements, the following search string was used in the automatic search:

(model OR method OR approach OR framework) AND (predict) AND (election*) AND ("social media" OR twitter OR facebook OR instagram)*.

The snowballing search on the references was applied only at the end of study selection to perform this search only on already identified relevant studies.

D. Study Selection

Study selection was performed by applying inclusion and exclusion criteria.

Inclusion criteria consisted of:

- I.1 – Articles written in English.
- I.2 – Articles published in peer-reviewed journals, in conferences or workshop proceedings, or conference papers published as book chapters.
- I.3 – Articles published as full papers.
- I.4 – Articles focused on predicting elections based on

SM data, or literature reviews or surveys of studies with this focus.

Exclusion criteria consisted of:

- E.1 – Short papers, tools session or demonstration papers, theses, technical reports, and books.
- E.2 – Duplicate studies.
- E.3 – Publications in which predicting elections based on SM data was not the focus
- E.4 – Studies not focused on electoral vote outcomes, such as only studying how politicians used SM during an electoral period.

This study included the identification of literature reviews and surveys (I.4) in order to have a more robust basis for the snowballing strategy. Also, they form a baseline of similar studies for us to compare our results with and to add issues and features not considered in previous studies.

The criteria used to verify short papers (I.3 and E.1), were (i) articles that were clearly mentioned as one of these categories; and (ii) articles with no more than 4 pages, including references. In terms of SM data, we considered papers related to at least one of the main worldwide, well-known popular SNs—Facebook, Twitter and Instagram—or that clearly mentioned it was based on SM, such as papers related to the major local SN in Taiwan, the PTT Bulletin Board System, and related to Nairand, a social network targeted at Nigerians.

For duplicate studies, a repetition of the same paper from the same authors may be gathered many times from different search databases. In this case, it is considered only once. Also, sometimes a conference paper is invited to be resubmitted to a journal as an extended version, but its core methodology and results are the same. If that is identified, only the journal version would be included.

For study selection, three steps were performed:

1) Quick Scanning:

An initial filtering was applied by reading paper titles and abstracts and removing papers clearly out of desired context. Also, papers either not passing on inclusion criteria (I.1 - I.3) or failing in exclusion criteria (E.1 and E.2) were summarily excluded.

2) Entirely reading:

The initially identified publications were decided by further reviewing the full text in order to identify if I.4 was accepted and E.3 and E.4 were rejected.

3) Snowballing selection

To further find possibly missed papers, the selected papers' references were also scanned. Then, the new papers passed through the same process as defined in the previous steps, in a single round.

E. Quality Assessment

One initial difficulty regarding the quality assessment is that there is no established manner with which to define study "quality." In this study we used the premise suggested by [37], in which quality relates to the extent to which the study minimizes bias and maximizes internal and external validity. Thus, we focused the quality assessment on the rigor of the study. Hence, we proposed the following quality assessment questions:

- QA1: Are the aim(s)/objective(s) clearly identified?
- QA2: Is the related work comprehensively reviewed?
- QA3: Are the findings/results clearly reported?
- QA4: Are bias and threats to validity clearly discussed?
- QA5: Did the study compare the proposed solution and results with other works?

The scoring procedure was: Yes, if the study clearly answered the question; Partially, if the answers were implicit or could be inferred by the reader, or was very briefly presented; or No, if the study did not address the question. In particular, QA4 was set as Partially if the discussion focused on explaining or justifying unexpected or challenged results. The quality score was Yes = 1, Partially = 0.5 and No = 0 for each question. The overall quality of a publication was calculated by adding all the quality scores received.

F. Data Extraction and Synthesis

In accordance with previously defined research questions, this study used a data extraction schema to collect relevant data from studies, as listed in Table XII The schema covers a set of attributes, and each attribute corresponds to a data extraction question. The relationships between the data extraction questions and predefined research questions are also specified.

TABLE XII DATA EXTRACTION SCHEMA

Category	Data Extraction Attribute	Data Extraction Question	Quest.
Demography	Author	Who is/are the author(s)?	(Metadata)
	Affiliation	What is/are the author's affiliation(s)?	
	Publication Title	What is the title of publication?	
	Publication Year	In which year was the work published?	
	Keywords	What is/are the publication's keywords?	
	Number of Pages	What is the publication's number of pages?	
	Venue Type	What is the venue type in which the study was reported? (J = Journal; CWS = Conference of Workshop)	
	Venue Name	What is the publication's venue? (Name and acronym)	
Quality Assessment	Objective	QA1: Are the aim(s)/objective(s) clearly identified?	QA1
	Related Works	QA2: Is the related work comprehensively reviewed?	QA2
	Findings and results	QA3: Are the findings/results clearly reported?	QA3
	Bias, threats to validity	QA4: Are bias and threats to validity clearly discussed?	QA4
	Study comparison	QA5: Did the study compare the proposed solution and results with other works?	QA5
Elections Data	Number of elections studied	In how many elections the approach was applied?	RQ1
	Number of subelections	In how many sub-elections the approach was applied?	RQ1
	Election Coverage	What is the election(s) coverage? (Municipal, State or National)	RQ1
	Election Role	What is the election(s) role? (presidential, assembly, etc.)	RQ1
	Type of vote	Who is voted for? (direct or party)	RQ1
	Election Year	What is the election's year?	RQ1
	Election Country	What is the election's country?	RQ1
	Number of candidates	How many candidates was running or analyzed?	RQ1
Model	Main Proposed Model	What is the main model addressed in the study?	RQ2
	Secundary Proposed models	What are the secundary models addressed in the study?	RQ2
	Input Data: Social Networks	What are the social network used as input data?	RQ2
	Input Data: Other datasources	What are the secundary datasources used as input data?	RQ2
Data Collection	Data Type	What was the main data type used in the study?	RQ2
	Type of collection	How was the approach for data collection?	RQ2
	Data Volume	How much data was collected?	RQ2
	Collection Time	When the data was collected?	RQ2
	Collection Duration	What was the lenght, in days, of data collection?	RQ2
Prediction	Prediction Type	What is the prediction type (continuos, one shot prediction, few shot predictions)	RQ2
	Prediction Time in theory	When, in theory, results was predicted (days before)	RQ2
	Prediction Time in the practice	Predictions was actually performed before elections?	RQ2
Metrics	Prediction Metric	What is the main metric for predicting? (vote share, winner, other)	RQ2
	Error Metrics	What is the main error metric for results evaluation?	RQ2
	Statistical Analysis	Was it applied statistical tests for results evaluation and validation?	RQ2
Results	Claimed results	Does the paper achieve positive results?	RQ3

In particular, the collected data is composed by publication metadata, quality assessment data, and evaluation data. The metadata was used to draw a general picture of research in the field and included demographic data such as authors' names and affiliations, publication title, keywords, year published, and number of pages, as well as the venue type and name. Quality assessment was extracted as already defined. Evaluation data was analyzed to answer the research questions, and includes data regarding the context of studied election, the model used for prediction, data collection approach, metrics used for performance evaluation and validation, and additional information regarding the results.

II. Study Selection

The search procedure was performed twice. The first execution was completed on July 31, 2019 and generated the first version of this study with incomplete data about the 2019 year. One year later, a new search was performed to gather the remaining papers published in 2019.

Considering the first search procedure, by applying the search string in the four indexing systems, 525 papers were gathered: 232 from Scopus, 154 from Web of Science, 71 from ACM, and 68 from IEEEXplorer. Then, the selection of studies was performed as described in Appendix I and is summarized in Fig. 5.

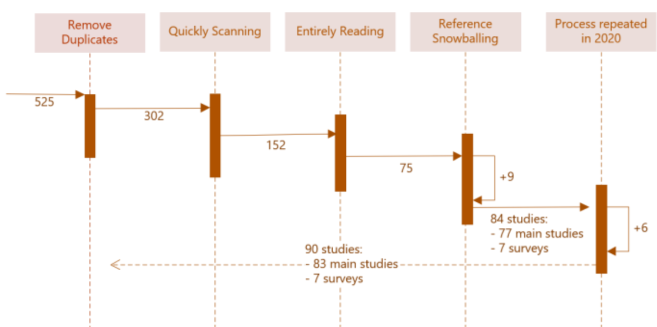

Fig. 5. Study selection phases

By removing duplicates, 302 papers were considered. In the quick scanning phase, papers clearly out of desired context were removed (e.g., "Predicting iPhone Sales from iPhone Tweets" and "Identifying Controversial Wikipedia Articles Using Editor Collaboration Networks"), and the first set of inclusion and exclusion criteria (I.1 – I.3 and E.1 – E.2) was applied, resulting in 152 papers. Then, after the entire reading phase, the last criteria (I.4 and E.3 – E.4), which consider whether the study focused on predicting elections based on social media data or if the study was a literature review or survey of studies with this focus. In this analysis, we admitted one exception: the work of Tilton [4] that uses SM to predict elections in a university was included due to its pioneering work. From this, a list of 75 studies was generated. Finally, by applying snowballing in the references of these papers, 9 new studies were found, analyzed, and selected, resulting in a final set of 84 studies: 77 main primary studies and 7 surveys or literature reviews.

The process was repeated one year after, focusing on gathering the remaining papers from the year 2019. As a result, six new primary studies were added, leading to a final set of 90 studies: 83 primary studies and 7 surveys or literature reviews. Primary studies were analyzed and discussed to answer the research questions, whilst surveys were used in the discussion and comparison of this works' results. The list of full references of all 83 primary studies is presented in Appendix III. The six new papers are at the end of list, from PS78 to PS83.

The authors would like to highlight that three studies were excluded despite containing expressions like "predicting election" on their titles. The papers "Predicting Election Results using NLTK", "Prediction of Indonesia Presidential Election Results for the 2019-2024 Period Using Twitter Sentiment Analysis", and "Sentiment Analysis of Twitter for Election Prediction" where excluded by criteria E.4, because they only present sentiment analysis or SM analysis without a relation with electoral prediction.

III. Studies Main Characteristics

TABLE XIII. STUDIES MAIN CHARACTERISTICS

ID	Elec.	Sub Elec.	Role	Vote	Year	Country	Develop.	Candidates	Model	Social Network	Data Volume	Collection Days	Pred. Freq.	Pred. Metric	Success
PS01	1	N	Parliam.	Party	2010	U.K.	Developed	3-5	RT;	FB; TT; OT; PO	NI	1 - 30	One	VS	Yes
PS02	1	N	Parliam.	Party	2018	Pakistan	Developing	3-5	VS;	TT	100k - 500k	NI	One	Other	No
PS03	1	N	Prim.	Direct	2012	U.S.	Developed	6-10	VS; PPI;	FB; TT; OT	NI	> 90	Many	VS	Yes
PS04	1	N	Prim.	Direct	2016	U.S.	Developed	1-2	VS;	TT	100k - 500k	31 - 90	One	VS	No
PS05	2	N	Mult.	Direct	2012; 2014	U.S.	Developed	Mult.	VS; RT;	OT; PO	NI	1 - 30	Many	VS	Yes
PS06	2	N	Mult.	Mult.	2012; 2013	Mult.	Mult.	Mult.	VS; PPI;	TT	< 100k	31 - 90	Many	VS	NC
PS07	1	N	Presid.	Direct	2016	U.S.	Developed	1-2	VS;	TT	< 100k	1 - 30	One	Winner	No
PS08	1	N	Presid.	Direct	2014	Colombia	Developing	3-5	VS; RT;	TT; PO	>= 500k	1 - 30	One	VS	NC
PS09	1	S	Prim.	Direct	2016	U.S.	Developed	>12	VS;	TT	< 100k	31 - 90	One	Other	No
PS10	1	N	Prim.	Direct	2012	U.S.	Developed	3-5	VS;	TT	>= 500k	31 - 90	One	VS	Yes
PS11	1	Y	Parliam.	Party	2015	Venezuela	Developing	1-2	OT	TT	< 100k	1 - 30	One	Winner	Yes
PS12	1	N	Presid.	Direct	2016	Taiwan	Developing	3-5	RT;	FB; TT; OT; PO	NI	31 - 90	Many	VS	Yes
PS13	1	Y	Other	Direct	2014	Taiwan	Developing	>10	VS; PPI;	OT	NI	> 90	Many	VS	NC
PS14	1	N	Presid.	Direct	2014	Indonesia	Developing	1-2	VS;	TT	>= 500k	31 - 90	One	VS	Yes
PS15	1	N	Other	Direct	2017	Indonesia	Developing	3-5	VS;	TT	< 100k	NI	One	Winner	No
PS16	1	N	Presid.	Direct	2008	U.S.	Developed	1-2	VS;	TT	100k - 500k	> 90	One	VS	No
PS17	1	N	Presid.	Direct	2016	U.S.	Developed	1-2	VS;	TT	100k - 500k	1 - 30	One	Winner	No
PS18	1	N	Presid.	Direct	2016	U.S.	Developed	1-2	VS; RT;	TT; OT; PO	100k - 500k	> 90	Many	VS	No
PS19	1	N	Prim.	Direct	2013	Italy	Developed	3-5	VS; RT;	TT	< 100k	1 - 30	One	VS	Yes
PS20	1	N	Presid.	Direct	2017	Iran	Developing	3-5	VS; PPI;	TT; OT	>= 500k	31 - 90	One	VS	Yes
PS21	1	N	Parliam.	Party	2012	France	Developed	6-10	VS;	TT	< 100k	1 - 30	One	VS	Yes
PS22	2	N	Mult.	Mult.	2016; 2017	Mult.	Developed	1-2	VS; TA;	TT	>= 500k	31 - 90	Many	VS	No
PS23	1	N	Presid.	Direct	2016	U.S.	Developed	1-2	VS;	TT	>= 500k	31 - 90	Many	VS	Yes
PS24	1	N	Parliam.	Direct	2015	Finland	Developed	>14	PPI;	FB	< 100k	1 - 30	One	VS	No
PS25	1	Y	Other	Direct	2012	U.S.	Developed	1-2	PPI;	FB	NI	31 - 90	Many	VS	Yes
PS26	1	N	Presid.	Direct	2016	U.S.	Developed	1-2	VS;	TT	100k - 500k	31 - 90	One	Winner	No
PS27	1	N	Presid.	Direct	2017	Chile	Developing	6-10	VS;	TT	>= 500k	> 90	Many	VS	NC
PS28	1	N	Presid.	Direct	2008	U.S.	Developed	1-2	VS; RT;	TT; PO	>= 500k	NI	Many	VS	No
PS29	1	Y	Other	Direct	2010	U.S.	Developed	1-2	VS;	TT	100k - 500k	1 - 30	One	VS	No
PS30	1	Y	Other	Direct	2011	Italy	Developed	6-10	PPI;	FB	< 100k	1 - 30	One	VS	No
PS31	1	N	Parliam.	Party	2014	India	Developing	3-5	PPI;	FB	>= 500k	> 90	One	Other	Yes
PS32	2	N	Mult.	Mult.	2012; 2013	Mult.	Mult.	Mult.	VS; PPI;	TT	100k - 500k	31 - 90	One	VS	NC
PS33	1	N	Parliam.	Party	2018	Pakistan	Developing	3-5	VS; PPI;	TT; PO	>= 500k	1 - 30	Many	Other	Yes
PS34	3	N	Presid.	Direct	2013	Mult.	Developing	Mult.	VS; RT;	TT	>= 500k	1 - 30	Many	VS	Yes
PS35	1	N	Parliam.	Party	2017	India	Developing	3-5	VS;	TT	100k - 500k	1 - 30	One	VS	No
PS36	1	Y	Other	Direct	2010	U.S.	Developed	1-2	VS;	TT	100k - 500k	1 - 30	One	VS	No
PS37	1	Y	Other	Direct	2014	Taiwan	Developing	1-2	PPI; TA;	OT; PO	NI	> 90	One	Winner	Yes
PS38	1	N	Parliam.	Party	2013	Pakistan	Developing	3-5	VS;	TT	< 100k	> 90	One	Winner	No
PS39	1	N	Parliam.	Party	2014	India	Developing	1-2	VS;	TT	< 100k	NI	One	VS	Yes
PS40	1	N	Parliam.	Party	2013	Bulgaria	Developed	3-5	VS;	TT	< 100k	1 - 30	One	VS	No
PS41	1	N	Parliam.	Party	2015	Venezuela	Developing	1-2	VS; TA;	TT	< 100k	31 - 90	One	VS	Yes
PS42	1	N	Other	Direct	2015	Greece	Developed	1-2	VS; PPI;	TT	>= 500k	1 - 30	Many	VS	Yes

Y = Yes, N = No, Parliam. = Parliament, Presid. = Presidential, Mult. = Multiple, VS = Volume or Sentiment, RT = Regression or Time Series, PPI = Profile or Posts Interactions, TA = Topic Analysis, OT = Other, TT = Twitter, FB = Facebook, PO = Polls, VS = Vote Share.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TCSS.2021.3063660

TABLE XIII. STUDIES MAIN CHARACTERISTICS (CONTINUED)

ID	Elec.	Sub Elec.	Role	Vote	Year	Country	Develop.	Candidates	Model	Social Network	Data Volume	Collection Days	Pred. Freq.	Pred. Metric	Success
PS43	1	N	Parliam.	Party	2010	U.K.	Developed	3-5	VS; RT;	TT; PO	100k - 500k	1 - 30	Many	VS	Yes
PS44	1	N	Parliam.	Party	2017	India	Developing	3-5	VS; TA;	TT	100k - 500k	1 - 30	One	VS	Yes
PS45	1	N	Parliam.	Party	2011	Ireland	Developed	3-5	VS; RT;	TT; PO	< 100k	1 - 30	Many	VS	No
PS46	1	N	Prim.	Direct	2016	U.S.	Developed	6-10	RT;	FB; TT; OT; PO	NI	31 - 90	Many	Other	Yes
PS47	1	Y	Presid.	Direct	2012	U.S.	Developed	1-2	RT;	TT; PO	>= 500k	31 - 90	Many	VS	Yes
PS48	3	N	Parliam.	Party	2014	Mult.	Developed	Mult.	RT;	TT; PO	100k - 500k	31 - 90	Many	VS	Yes
PS49	3	N	Parliam.	Party	2013; 2014	Mult.	Developing	Mult.	VS; OT	TT	>= 500k	NI	One	VS	Yes
PS50	1	N	Parliam.	Party	2013	Australia	Developed	1-2	VS; TA;	TT	>= 500k	31 - 90	One	VS	Yes
PS51	1	N	Parliam.	Party	2009	Germany	Developed	6-10	VS;	TT	100k - 500k	31 - 90	One	VS	Yes
PS52	1	N	Other	Party	2011	The Neth.	Developed	>11	VS;	TT	< 100k	1 - 30	One	VS	No
PS53	1	N	Parliam.	Party	2016	Spain	Developed	3-5	VS;	TT	< 100k	1 - 30	One	VS	No
PS54	1	N	Presid.	Direct	2019	Indonesia	Developing	1-2	VS;	TT	< 100k	> 90	One	Winner	Yes
PS55	1	N	Parliam.	Party	2016	India	Developing	3-5	VS;	TT	< 100k	1 - 30	One	Winner	Yes
PS56	1	N	Presid.	Direct	2017	France	Developed	1-2	VS;	TT	NI	1 - 30	Many	VS	Yes
PS57	1	N	Presid.	Direct	2017	France	Developed	1-2	VS;	TT	NI	1 - 30	One	VS	Yes
PS58	1	Y	Prim.	Direct	2012	U.S.	Developed	6-10	PPI; OT	TT; PO	>= 500k	1 - 30	One	VS	Yes
PS59	1	N	Parliam.	Party	2010	U.K.	Developed	3-5	VS;	TT	>= 500k	31 - 90	One	VS	Yes
PS60	1	N	Presid.	Direct	2016	U.S.	Developed	1-2	VS;	TT	100k - 500k	> 90	One	Winner	Yes
PS61	1	N	Presid.	Direct	2016	U.S.	Developed	1-2	VS;	TT	>= 500k	31 - 90	Many	VS	NC
PS62	1	Y	Mult.	Direct	2016	Taiwan	Developing	1-2	RT;	FB; OT; PO	< 100k	NI	One	VS	Yes
PS63	1	N	Prim.	Direct	2017	Chile	Developing	3-5	VS;	TT	>= 500k	31 - 90	Many	VS	Yes
PS64	1	N	Presid.	Direct	2014	Brazil	Developing	1-2	VS;	TT	100k - 500k	1 - 30	Many	Winner	Yes
PS65	1	Y	Other	Direct	2011	Italy	Developed	1-2	OT	FB	NI	NI	One	Winner	Yes
PS66	1	N	Parliam.	Party	2015	India	Developing	3-5	VS;	TT; PO	>= 500k	31 - 90	One	VS	Yes
PS67	2	N	Mult.	Direct	2014; 2016	U.S.	Developed	1-2	VS; RT;	TT	NI	1 - 30	Many	Other	Yes
PS68	3	N	Parliam.	Party	2011; 2012	Spain	Developed	Mult.	VS;	TT	100k - 500k	1 - 30	Many	VS	Yes
PS69	1	N	Presid.	Direct	2012	U.S.	Developed	6-10	VS;	TT	>= 500k	1 - 30	Many	VS	Yes
PS70	1	N	Other	Direct	2017	Indonesia	Developing	3-5	VS;	TT	100k - 500k	31 - 90	One	VS	NC
PS71	1	N	Presid.	Direct	2014	Indonesia	Developing	1-2	VS;	TT	>= 500k	31 - 90	One	VS	Yes
PS72	2	N	Presid.	Direct	2012	Mult.	Developed	Mult.	VS; RT;	TT	NI	31 - 90	Many	VS	Yes
PS73	1	N	Presid.	Direct	2016	U.S.	Developed	1-2	OT	TT	>= 500k	> 90	Many	VS	Yes
PS74	1	N	Other	Direct	2008	U.S.	Developed	>13	OT	FB	< 100k	1 - 30	One	VS	Yes
PS75	1	N	Other	Direct	2014	Taiwan	Developing	1-2	TA;	FB; OT	NI	NI	One	VS	Yes
PS76	1	N	Parliam.	Party	2009	Germany	Developed	6-10	VS;	TT	100k - 500k	31 - 90	One	VS	No
PS77	1	N	Presid.	Direct	2016	Taiwan	Developing	3-5	RT;	FB; TT; OT; PO	100k - 500k	> 90	Many	VS	Yes
PS78	1	N	Parliam.	Party	2015	India	Developing	1-2	VS;	TT	>= 500k	> 90	One	Winner	Yes
PS79	1	N	Parliam.	Party	2018	Pakistan	Developing	3-5	PPI;	FB	< 100k	> 90	One	VS	Yes
PS80	1	N	Parliam.	Party	2017	India	Developing	3-5	VS;	TT	< 100k	1 - 30	One	VS	Yes
PS81	3	N	Parliam.	Party	2013; 2014; 2016	Japan	Developed	Mult.	VS; RT;	FB; TT	>= 500k	1 - 30	One	VS	NC
PS82	1	N	Presid.	Direct	2019	Nigeria	Developing	1-2	VS;	TT	>= 500k	31 - 90	One	VS	No
PS83	1	N	Presid.	Direct	2019	Nigeria	Developing	1-2	VS;	OT	100k - 500k	31 - 90	One	Winner	Yes

Y = Yes, N = No, Parliam. = Parliament, Presid. = Presidential, Mult. = Multiple, VS = Volume or Sentiment, RT = Regression or Time Series, PPI = Profile or Posts Interactions, TA = Topic Analysis, OT = Other, TT = Twitter, FB = Facebook, PO = Polls, VS = Vote Share.

IV. LIST OF PRIMARY STUDIES

[PS01] F. Franch, "(Wisdom of the Crowds) 2 : 2010 UK Election Prediction with Social Media," J. Inf. Technol. Polit., vol. 10, no. 1, pp. 57–71, Jan. 2013.

[PS02] M. Bilal, S. Asif, S. Yousuf, and U. Afzal, "2018 Pakistan General Election: Understanding the Predictive Power of Social Media," in 12th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), 2018, pp. 1–6.

[PS03] F. Pimenta, D. Obradović, and A. Dengel, "A comparative study of social media prediction potential in the 2012 U.S. Republican presidential preelections," in Proceedings - 2013 IEEE 3rd International Conference on Cloud and Green Computing, CGC 2013 and 2013 IEEE 3rd International Conference on Social Computing and Its Applications, SCA 2013, 2013.

[PS04] L. Oikonomou and C. Tjortjis, "A Method for Predicting the Winner of the USA Presidential Elections using Data extracted from Twitter," in 2018 South-Eastern European Design Automation, Computer Engineering, Computer Networks and Society Media Conference (SEEDA_CECNSM), 2018, pp. 1–8.

[PS05] Q. You, L. Cao, Y. Cong, X. Zhang, and J. Luo, "A Multifaceted Approach to Social Multimedia-Based Prediction of Elections," IEEE Trans. Multimed., vol. 17, no. 12, pp. 2271–2280, Dec. 2015.

[PS06] M. Anjaria and R. M. R. Guddeti, "A novel sentiment analysis of social networks using supervised learning," Soc. Netw. Anal. Min., vol. 4, no. 1, p. 181, Dec. 2014.

[PS07] Andy Januar Wicaksono, Suyoto, and Pranowo, "A proposed method for predicting US presidential election by analyzing sentiment in social media," in 2016 2nd International Conference on Science in Information Technology (ICSITech), 2016, pp. 276–280.

[PS08] J. A. Ceron-Guzman and E. Leon-Guzman, "A Sentiment Analysis System of Spanish Tweets and Its Application in Colombia 2014 Presidential Election," in 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom), 2016, pp. 250–257.

[PS09] E. Tunggawan and Y. E. Soelistio, "And the winner is...: Bayesian Twitter-based prediction on 2016 U.S. presidential election," in Proceeding - 2016 International Conference on Computer, Control, Informatics and its Applications: Recent Progress in Computer, Control, and Informatics for Data Science, IC3INA 2016, 2017.

[PS10] L. Chen, W. Wang, and A. P. Sheth, "Are twitter users equal in predicting elections? A study of user groups in predicting 2012 U.S. republican presidential primaries," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012.

[PS11] R. Castro, L. Kuffó, and C. Vaca, "Back to #6D: Predicting Venezuelan states political election results through Twitter," in 2017 4th International Conference on eDemocracy and eGovernment, ICEDEG 2017, 2017.

[PS12] Z. Xie, G. Liu, J. Wu, and Y. Tan, "Big data would not lie: prediction of the 2016 Taiwan election via online heterogeneous information," EPJ Data Sci., 2018.

[PS13] M. H. Wang and C. L. Lei, "Boosting election prediction accuracy by crowd wisdom on social forums," in 2016 13th IEEE Annual Consumer Communications and Networking Conference, CCNC 2016, 2016.

[PS14] M. Ibrahim, O. Abdillah, A. F. Wicaksono, and M. Adriani, "Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in a Twitter Nation," in 2015 IEEE International Conference on Data Mining Workshop (ICDMW), 2015, pp. 1348–1353.

[PS15] D. A. Kristiyanti, A. H. Umam, M. Wahyudi, R. Amin, and L. Marlinda, "Comparison of SVM Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter," in 6th International Conference on Cyber and IT Service Management, CITSM 2018, 2019.

[PS16] D. Gayo-Avello, "Don't turn social media into another 'Literary Digest' poll," Commun. ACM, vol. 54, no. 10, p. 121, Oct. 2011.

[PS17] J. Ramteke, S. Shah, D. Godhia, and A. Shaikh, "Election result prediction using Twitter sentiment analysis," in 2016 International Conference on Inventive Computation Technologies (ICICT), 2016, pp. 1–5.

[PS18] P. Kassraie, A. Modirshanechi, and H. K. Aghajan, "Election Vote Share Prediction using a Sentiment-based Fusion of Twitter Data with Google Trends and Online Polls," in Proceedings of the 6th International Conference on Data Science, Technology and Applications, 2017, pp. 363–370.

[PS19] M. Coletto, C. Lucchese, S. Orlando, and R. Perego, "Electoral predictions with Twitter: A machine-learning approach," in CEUR Workshop Proceedings, 2015.

[PS20] S. Salari, N. Sedighpour, V. Vaezinia, and S. Momtazi, "Estimation of 2017 Iran's Presidential Election Using Sentiment Analysis on Social Media," in 2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), 2018, pp. 77–82.

[PS21] A. Ceron, L. Curini, S. M. Iacus, and G. Porro, "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 Soc., vol. 16, no. 2, pp. 340–358, Mar. 2014.

[PS22] X. Hu, L. Li, T. Wu, X. Ai, J. Gu, and S. Wen, "Every word is valuable: Studied influence of negative words that spread during election period in social media," Concurr. Comput. Pract. Exp., vol. 31, no. 21, Nov. 2019.

[PS23] B. Heredia, J. Prusa, and T. Khoshgoftaar, "Exploring the Effectiveness of Twitter at Polling the United States 2016 Presidential Election," in Proceedings - 2017 IEEE 3rd International Conference on Collaboration and Internet Computing, CIC 2017, 2017.

[PS24] T. Vepsäläinen, H. Li, and R. Suomi, "Facebook likes and public opinion: Predicting the 2015 Finnish parliamentary elections," Gov. Inf. Q., vol. 34, no. 3, pp. 524–532, Sep. 2017.

[PS25] M. C. MacWilliams, "Forecasting Congressional Elections Using Facebook Data," PS Polit. Sci. Polit., vol. 48, no. 04, pp. 579–583, Oct. 2015.

[PS26] P. Singh, R. S. Sawhney, and K. S. Kahlon, "Forecasting the 2016 US presidential elections using sentiment analysis," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017.

[PS27] S. Rodríguez et al., "Forecasting the Chilean Electoral Year: Using Twitter to Predict the Presidential Elections of 2017," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, pp. 298–314.

[PS28] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series," in 4th International AAAI Conference on Weblogs and Social Media, 2010.

[PS29] P. T. Metaxas, E. Mustafaraj, and D. Gayo-Avello, "How (Not) to predict elections," in Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011, 2011.

[PS30] F. Giglietto, "If Likes were votes: An empirical study on the 2011 Italian administrative elections," in ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media, 2012.

[PS31] F. P. Barclay, C. Pichandy, A. Venkat, and S. Sudhakaran, "India 2014: Facebook 'Like' as a Predictor of Election Outcomes," Asian J. Polit. Sci., vol. 23, no. 2, pp. 134–160, May 2015.

[PS32] M. Anjaria and R. M. R. Guddeti, "Influence factor based opinion mining of Twitter data using supervised learning," in 2014 Sixth International Conference on Communication Systems and Networks (COMSNETS), 2014, pp. 1–8.

[PS33] M. Awais, S.-U. Hassan, and A. Ahmed, "Leveraging big data for politics: predicting general election of Pakistan using a novel rigged model," J. Ambient Intell. Humaniz. Comput., Jul. 2019.

[PS34] M. Gaurav, A. Srivastava, A. Kumar, and S. Miller, "Leveraging candidate popularity on Twitter to predict election outcome," in Proceedings of the 7th Workshop on Social Network Mining and Analysis - SNAKDD '13, 2013, pp. 1–8.

[PS35] B. Bansal and S. Srivastava, "Lexicon-based Twitter sentiment analysis for vote share prediction using emoji and N-gram features," Int. J. Web Based Communities, vol. 15, no. 1, p. 85, 2019.

[PS36] D. Gayo-Avello, P. T. Metaxas, and E. Mustafaraj, "Limits of Electoral Predictions using Social Media Data," Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011.

[PS37] K.-C. Tung, E. T. Wang, and A. L. P. Chen, "Mining Event Sequences from Social Media for Election Prediction," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016, pp. 266–281.

[PS38] T. Mahmood, T. Iqbal, F. Amin, W. Lohanna, and A. Mustafa, "Mining Twitter big data to predict 2013 Pakistan election winner," in Proceedings of International Multi Topic Conference, 2013, pp. 49–54.

[PS39] K. Singhal, B. Agrawal, and N. Mittal, "Modeling indian general elections: Sentiment analysis of political twitter data," in Advances in Intelligent Systems and Computing, 2015.

[PS40] J. Smailović, J. Kranjc, M. Grčar, M. Žnidaršič, and I. Mozetič, "Monitoring the Twitter sentiment during the Bulgarian elections," in Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015, 2015.

[PS41] R. Castro and C. Vaca, "National Leaders' Twitter Speech to Infer Political Leaning and Election Results in 2015 Venezuelan Parliamentary Elections," in 2017 IEEE International Conference on Data Mining Workshops (ICDMW), 2017, pp. 866–871.

[PS42] A. Tsakalidis, N. Aletras, A. I. Cristea, and M. Liakata, "Nowcasting the Stance of Social Media Users in a Sudden Vote: The case of the Greek Referendum," in Proceedings of the 27th ACM International Conference on Information and Knowledge Management - CIKM '18, 2018, pp. 367–376.

[PS43] E. Kalampokis, A. Karamanou, E. Tambouris, and K. Tarabanis, "On predicting election results using twitter and linked open data: The case of the UK 2010 election," J. Univers. Comput. Sci., vol. 23, no. 3, 2017.

[PS44] B. Bansal and S. Srivastava, "On predicting elections with hybrid topic based sentiment analysis of tweets," in Procedia Computer Science, 2018.

[PS45] A. Bermingham and A. F. Smeaton, "On Using Twitter to Monitor Political Sentiment and Predict Election Results," in Workshop on Sentiment Analysis where AI meets Psychology, 2011, pp. 2–10.

[PS46] V. Isotalo, P. Saari, M. Paasivaara, A. Steineker, and P. A. Gloor, "Predicting 2016 US presidential election polls with online and media variables," in Springer Proceedings in Complexity, 2016.

[PS47] N. Beauchamp, "Predicting and Interpolating State-Level Polls Using Twitter Textual Data," Am. J. Pol. Sci., vol. 61, no. 2, pp. 490–503, Apr. 2017.

[PS48] A. Tsakalidis, S. Papadopoulos, A. I. Cristea, and Y. Kompatsiaris, "Predicting Elections for Multiple Countries Using Twitter and Polls," IEEE Intell. Syst., 2015.
[PS49] K. Jaidka, S. Ahmed, M. Skoric, and M. Hilbert, "Predicting elections from social media: a three-country, three-method comparative study," Asian J. Commun., vol. 29, no. 3, pp. 252–273, May 2019.
[PS50] S. Unankard, X. Li, M. Sharaf, J. Zhong, and X. Li, "Predicting Elections from Social Networks Based on Sub-event Detection and Sentiment Analysis," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2014, pp. 1–16.
[PS51] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," in Fourth International AAAI Conference on Weblogs and Social Media, 2010.
[PS52] E. Sang and J. Bos, "Predicting the 2011 dutch senate election results with twitter," in Proceedings of the Workshop on Semantic Analysis in Social Media, 2012, pp. 53–60.
[PS53] P. Singh, R. S. Sawhney, and K. S. Kahlon, "Predicting the Outcome of Spanish General Elections 2016 Using Twitter as a Tool," in Communications in Computer and Information Science, 2017, pp. 73–83.
[PS54] W. Budiharto and M. Meiliana, "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis," J. Big Data, vol. 5, no. 1, p. 51, Dec. 2018.
[PS55] P. Sharma and T.-S. Moh, "Prediction of Indian election using sentiment analysis on Hindi Twitter," in 2016 IEEE International Conference on Big Data (Big Data), 2016, pp. 1966–1971.
[PS56] L. Wang and J. Q. Gan, "Prediction of the 2017 French election based on Twitter data analysis," in 2017 9th Computer Science and Electronic Engineering Conference, CEEC 2017 - Proceedings, 2017.
[PS57] L. Wang and J. Q. Gan, "Prediction of the 2017 French Election Based on Twitter Data Analysis Using Term Weighting," in 2018 10th Computer Science and Electronic Engineering (CEEC), 2018, pp. 231–235.
[PS58] M. J. Jensen and N. Anstead, "Psephological investigations: Tweets, votes, and unknown unknowns in the republican nomination process," in Policy and Internet, 2013.
[PS59] Y. He, H. Saif, Z. Wei, and K. F. Wong, "Quantising opinions for political tweets analysis," in Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012, 2012.
[PS60] G. Hu, S. Kodali, and A. Padamati, "Sentiment Analysis of Tweets on 2016 US Presidential Election Candidates," in 29th International Conference on Computer Applications in Industry and Engineering, 2016, p. 8.
[PS61] B. Heredia, J. D. Prusa, and T. M. Khoshgoftaar, "Social media for polling and predicting United States election outcome," Soc. Netw. Anal. Min., vol. 8, no. 1, p. 48, Dec. 2018.
[PS62] X. Zhang, "Social media popularity and election results: A study of the 2016 Taiwanese general election," PLoS One, 2018.
[PS63] P. Santander, C. Elórtegui, C. González, H. Allende-Cid, and W. Palma, "Social networks, computational intelligence and electoral prediction: the case of the presidential primaries of Chile 2017," Cuad. Inf., vol. 41, pp. 41–56, Apr. 2018.
[PS64] B. Justino Garcia Praciano, J. P. Carvalho Lustosa Da Costa, J. P. Abreu Maranhao, F. L. Lopes De Mendonca, R. T. De Sousa Junior, and J. Barbosa Prettz, "Spatio-Temporal trend analysis of the brazilian elections based on twitter data," in IEEE International Conference on Data Mining Workshops, ICDMW, 2019.
[PS65] P. Ceravolo and S. Guerretti, "Testing Social Network Metrics for Measuring Electoral Success in the Italian Municipal Campaign of 2011," in 2013 International Conference on Cloud and Green Computing, 2013, pp. 342–347.
[PS66] V. K. Jain and S. Kumar, "Towards Prediction of Election Outcomes Using Social Media," Int. J. Intell. Syst. Appl., vol. 9, no. 12, pp. 20–28, Dec. 2017.
[PS67] D. Watts, K. M. George, T. K. A. Kumar, and Z. Arora, "Tweet sentiment as proxy for political campaign momentum," in 2016 IEEE International Conference on Big Data (Big Data), 2016, pp. 2475–2484.
[PS68] J. M. Soler, F. Cuartero, and M. Roblizo, "Twitter as a Tool for Predicting Elections Results," in 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012, pp. 1194–1200.
[PS69] M. Maldonado and V. Sierra, "Twitter predicting the 2012 US presidential election?: Lessons learned from an unconscious value co-creation platform," J. Organ. End User Comput., 2016.
[PS70] D. F. Budiono, A. S. Nugroho, and A. Doewes, "Twitter sentiment analysis of DKI Jakarta's gubernatorial election 2017 with predictive and descriptive approaches," in 2017 International Conference on Computer, Control, Informatics and its Applications (IC3INA), 2017, pp. 89–94.
[PS71] N. D. Prasetyo and C. Hauff, "Twitter-based Election Prediction in the Developing World," Proc. 26th ACM Conf. Hypertext Soc. Media - HT '15, 2015.
[PS72] A. Ceron, L. Curini, and S. M. Iacus, "Using Sentiment Analysis to Monitor Electoral Campaigns: Method Matters—Evidence From the United States and Italy," Soc. Sci. Comput. Rev., vol. 33, no. 1, pp. 3–20, Feb. 2015.
[PS73] A. Bovet, F. Morone, and H. A. Makse, "Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump," Sci. Rep., vol. 8, no. 8673, Dec. 2018.
[PS74] S. Tilton, "Virtual polling data: A social network analysis on a student government election," Webology, 2008.
[PS75] J.-Y. Huang, "Web mining for the mayoral election prediction in Taiwan," Aslib J. Inf. Manag., vol. 69, no. 6, pp. 688–701, Nov. 2017.
[PS76] A. Jungherr, P. Jürgens, and H. Schoen, "Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, A., sprenger, T. O., sander, P. G., & welpe, I. M. 'predicting elections with twitter: What 140 characters reveal about political sentiment,'" Soc. Sci. Comput. Rev., 2012.
[PS77] Zheng Xie, Guannan Liu, Junjie Wu, Lihong Wang, and Chunyang Liu, "Wisdom of fusion: Prediction of 2016 Taiwan election with heterogeneous big data," in 2016 13th International Conference on Service Systems and Service Management (ICSSSM), 2016, pp. 1–6.
[PS78] L. Mohan and S. Elayidom, "Predicting the Winner of Delhi Assembly Election, 2015 from Sentiment Analysis on Twitter Data-A BigData Perspective," Int. Arab J. Inf. Technol., vol. 16, no. 5, pp. 833–842, 2019.
[PS79] M. Bilal, N. Malik, N. Bashir, M. Marjani, I. A. T. Hashem, and A. Gani, "Profiling Social Media Campaigns and Political Influence: The Case of Pakistani Politics," in 2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), 2019, pp. 1–7.
[PS80] Y. Gupta and P. Kumar, "Real-Time Sentiment Analysis of Tweets: A Case Study of Punjab Elections," in 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2019, pp. 1–12.
[PS81] H. Goto and Y. Goto, "Regression Analysis of National Elections in Japan Using Social Listening," in 12th IADIS International Conference Information Systems 2019, 2019, pp. 189–196.
[PS82] O. Oyebode and R. Orji, "Social Media and Sentiment Analysis: The Nigeria Presidential Election 2019," in 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2019, pp. 0140–0146.
[PS83] B. S. Bello, I. Inuwa-Dutse, and R. Heckel, "Social Media Campaign Strategies: Analysis of the 2019 Nigerian Elections," in 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), 2019, pp. 142–149.

References

[1] A. Jungherr, "Twitter use in election campaigns: A systematic literature review," *Journal of Information Technology and Politics*. 2016.
[2] P. L. Francia, "Free Media and Twitter in the 2016 Presidential Election: The Unconventional Campaign of Donald Trump," *Soc. Sci. Comput. Rev.*, vol. 36, no. 4, pp. 440–455, 2018.
[3] K. Brito, N. Paula, M. Fernandes, and S. Meira, "Social Media and Presidential Campaigns – Preliminary Results of the 2018 Brazilian Presidential Election," in *20th Annual International Conference on Digital Government Research on - dg.o 2019*, 2019, pp. 332–341.
[4] S. Tilton, "Virtual polling data: A social network analysis on a student government election," *Webology*, 2008.
[5] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," in *Fourth International AAAI Conference on Weblogs and Social Media*, 2010.
[6] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series," in *4th International AAAI Conference on Weblogs and Social Media*, 2010.
[7] E. Sang and J. Bos, "Predicting the 2011 dutch senate election results with twitter," in *Proceedings of the Workshop on Semantic Analysis in Social Media*, 2012, pp. 53–60.

[8] A. Ceron, L. Curini, S. M. Iacus, and G. Porro, "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 Soc.*, vol. 16, no. 2, pp. 340–358, Mar. 2014.

[9] K. Singhal, B. Agrawal, and N. Mittal, "Modeling indian general elections: Sentiment analysis of political twitter data," in *Advances in Intelligent Systems and Computing*, 2015.

[10] N. D. Prasetyo and C. Hauff, "Twitter-based Election Prediction in the Developing World," *Proc. 26th ACM Conf. Hypertext Soc. Media - HT '15*, 2015.

[11] J. A. Ceron-Guzman and E. Leon-Guzman, "A Sentiment Analysis System of Spanish Tweets and Its Application in Colombia 2014 Presidential Election," in *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)*, 2016, pp. 250–257.

[12] S. Rodríguez *et al.*, "Forecasting the Chilean Electoral Year: Using Twitter to Predict the Presidential Elections of 2017," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, pp. 298–314.

[13] O. Oyebode and R. Orji, "Social Media and Sentiment Analysis: The Nigeria Presidential Election 2019," in *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2019, pp. 0140–0146.

[14] B. S. Bello, I. Inuwa-Dutse, and R. Heckel, "Social Media Campaign Strategies: Analysis of the 2019 Nigerian Elections," in *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2019, pp. 142–149.

[15] B. Heredia, J. D. Prusa, and T. M. Khoshgoftaar, "Social media for polling and predicting United States election outcome," *Soc. Netw. Anal. Min.*, vol. 8, no. 1, p. 48, Dec. 2018.

[16] P. Singh, R. S. Sawhney, and K. S. Kahlon, "Forecasting the 2016 US presidential elections using sentiment analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017.

[17] A. Bovet, F. Morone, and H. A. Makse, "Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump," *Sci. Rep.*, vol. 8, no. 8673, Dec. 2018.

[18] M. Anjaria and R. M. R. Guddeti, "A novel sentiment analysis of social networks using supervised learning," *Soc. Netw. Anal. Min.*, vol. 4, no. 1, p. 181, Dec. 2014.

[19] D. Gayo-Avello, P. T. Metaxas, and E. Mustafaraj, "Limits of Electoral Predictions using Social Media Data," *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. 2011.

[20] D. Gayo-Avello, "Don't turn social media into another 'Literary Digest' poll," *Commun. ACM*, vol. 54, no. 10, p. 121, Oct. 2011.

[21] A. Jungherr, H. Schoen, O. Posegga, and P. Jürgens, "Digital Trace Data in the Study of Public Opinion," *Soc. Sci. Comput. Rev.*, vol. 35, no. 3, pp. 336–356, Jun. 2017.

[22] A. Jungherr, P. Jürgens, and H. Schoen, "Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, A., sprenger, T. O., sander, P. G., & welpe, I. M. 'predicting elections with twitter: What 140 characters reveal about political sentiment,'" *Soc. Sci. Comput. Rev.*, 2012.

[23] B. Bansal and S. Srivastava, "On predicting elections with hybrid topic based sentiment analysis of tweets," in *Procedia Computer Science*, 2018.

[24] S. Salari, N. Sedighpour, V. Vaezinia, and S. Momtazi, "Estimation of 2017 Iran's Presidential Election Using Sentiment Analysis on Social Media," in *2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, 2018, pp. 77–82.

[25] Andy Januar Wicaksono, Suyoto, and Pranowo, "A proposed method for predicting US presidential election by analyzing sentiment in social media," in *2016 2nd International Conference on Science in Information Technology (ICSITech)*, 2016, pp. 276–280.

[26] A. Tsakalidis, S. Papadopoulos, A. I. Cristea, and Y. Kompatsiaris, "Predicting Elections for Multiple Countries Using Twitter and Polls," *IEEE Intell. Syst.*, 2015.

[27] X. Zhang, "Social media popularity and election results: A study of the 2016 Taiwanese general election," *PLoS One*, 2018.

[28] V. Isotalo, P. Saari, M. Paasivaara, A. Steineker, and P. A. Gloor, "Predicting 2016 US presidential election polls with online and media variables," in *Springer Proceedings in Complexity*, 2016.

[29] E. Kalampokis, E. Tambouris, and K. Tarabanis, "Understanding the predictive power of social media," *Internet Res.*, vol. 23, no. 5, pp. 544–559, Oct. 2013.

[30] D. Gayo-Avello, "A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data," *Soc. Sci. Comput. Rev.*, vol. 31, no. 6, pp. 649–679, Dec. 2013.

[31] J. Prada, "Predicting With Twitter," in *European Conference on Social Media*, 2015, p. 10.

[32] D. E. O'Leary, "Twitter Mining for Discovery, Prediction and Causality: Applications and Methodologies," *Intell. Syst. Accounting, Financ. Manag.*, vol. 22, no. 3, pp. 227–247, Jul. 2015.

[33] J. Kwak and S. K. Cho, "Analyzing Public Opinion with Social Media Data during Election Periods: A Selective Literature Review," *Asian J. Public Opin. Res.*, vol. 5, no. 4, pp. 285–301, 2018.

[34] A. M. Koli, M. Ahmed, and J. Manhas, "An Empirical Study on Potential and Risks of Twitter Data for Predicting Election Outcomes," in *Advances in Intelligent Systems and Computing*, V. Rathore, M. Worring, D. Mishra, A. Joshi, and S. Maheshwari, Eds. Springer, Singapore, 2019, pp. 725–731.

[35] M. Bilal, A. Gani, M. Marjani, and N. Malik, "Predicting Elections: Social Media Data and Techniques," in *2019 International Conference on Engineering and Emerging Technologies (ICEET)*, 2019, pp. 1–6.

[36] We are Social and Hootsuite, "DIGITAL 2020: JULY GLOBAL STATSHOT," 2020. [Online]. Available: https://datareportal.com/reports/digital-2020-july-global-statshot. [Accessed: 18-Nov-2020].

[37] Centre for Reviews and Dissemination, *Systematic reviews: CRD's guidance for undertaking reviews in health care*. CRD, University of York,York, UK, 2009.

[38] H. Zhang, M. A. Babar, and P. Tell, "Identifying relevant studies in software engineering," *Inf. Softw. Technol.*, vol. 53, no. 6, pp. 625–637, Jun. 2011.

[39] F. Q. B. da Silva, A. L. M. Santos, S. Soares, A. C. C. França, C. V. F. Monteiro, and F. F. Maciel, "Six years of systematic literature reviews in software engineering: An updated tertiary study," *Inf. Softw. Technol.*, vol. 53, no. 9, pp. 899–913, Sep. 2011.

[40] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Inf. Softw. Technol.*, vol. 55, no. 12, pp. 2049–2075, Dec. 2013.

[41] Twitter Inc., "Twitter API," 2020. [Online]. Available: https://developer.twitter.com/en/docs. [Accessed: 21-Nov-2020].

[42] P. Langley, W. Iba, and K. Thompson, "An analysis of Bayesian classifiers," in *Proceedings of the Tenth National Conference on Artificial Intelligence*, 1992, pp. 223--228.

[43] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[44] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis*. Wiley, 2003.

[45] W. W. S. Wei, *Time Series Analysis*. Oxford University Press, 2013.

[46] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, no. Jan, pp. 993–1022, 2003.

[47] W. Jennings and C. Wlezien, "Election polling errors across time and space," *Nat. Hum. Behav.*, vol. 2, no. 4, pp. 276–283, Apr. 2018.

[48] United Nations, "World Economic Situation and Prospects 2018," 2018.

[49] Pew Research Center, "Sizing Up Twitter Users," 2019.

[50] Facebook Inc., "Facebook API," 2020. [Online]. Available: https://developers.facebook.com/docs/graph-api. [Accessed: 21-Nov-2020].

[51] T. Filer and R. Fredheim, "Popular with the Robots: Accusation and Automation in the Argentine Presidential Elections, 2015," *Int. J. Polit. Cult. Soc.*, vol. 30, no. 3, pp. 259–274, 2017.

[52] A. Bessi and E. Ferrara, "Social bots distort the 2016 U.S. Presidential election online discussion," *First Monday*, vol. 21, no. 11, Nov. 2016.

[53] E. Mustafaraj, S. Finn, C. Whitlock, and P. T. Metaxas, "Vocal minority versus silent majority: Discovering the opionions of the long tail," in *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, 2011.

[54] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explor.*, vol. 11, no. 1, 2009.

[55] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 85, p. 2825−2830, 2011.

[56] C. Shearer, "The CRISP-DM model: the new blueprint for data mining," *J. data Warehous.*, vol. 5, no. 4, pp. 13--22, 2000.

[57] M. Rogalewicz and R. Sika, "Methodologies of Knowledge Discovery from Data and Data Mining Methods in Mechanical Engineering," *Manag. Prod. Eng. Rev.*, vol. 7, no. 4, pp. 97–108, Dec. 2016.

[58] M. Alnoukari and A. El Sheikh, “Knowledge Discovery Process Models: : From Traditional to Agile Modeling,” in *Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications*, 2012, pp. 72–100.
[59] D. Kreiss, R. G. Lawrence, and S. C. McGregor, “In Their Own Words: Political Practitioner Accounts of Candidates, Audiences, Affordances, Genres, and Timing in Strategic Social Media Use,” *Polit. Commun.*, 2018.
[60] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Networks*, vol. 2, no. 5, pp. 359–366, Jan. 1989.
[61] N. Beauchamp, “Predicting and Interpolating State-Level Polls Using Twitter Textual Data,” *Am. J. Pol. Sci.*, vol. 61, no. 2, pp. 490–503, Apr. 2017.

Kellyton dos Santos Brito (M’15) received the B.S. degree in computer science from Universidade Federal da Bahia, Salvador, Brazil, in 2000 and the M.S. degree in computer science from Universidade Federal de Pernambuco, Recife, Brazil, in 2007. He is currently pursuing the Ph.D. degree in computer science at Universidade Federal de Pernambuco.

From 2008 to 2010 he was an Assistant Professor with the Universidade Estadual de Feira de Santana, Brazil, and from 2010 to 2011 he was an Assistant Professor with Universidade Federal de Alagoas. Since 2011, he has been an Adjunct Professor with Universidade Federal Rural de Pernambuco, Recife, Brazil. Besides, he is a researcher at INES (National Institute of Science and Technology in Software Engineering) and he carried out industrial projects in many companies, such as CESAR (Center for Advanced Studies and Systems of Recife), Brazil, and Medicware Systems, Brazil. His current research interest includes machine learning, elections forecasting, social media, open government data and software architecture.

Rogério Luiz Cardoso Silva Filho received the B.S. degree in information systems from Universidade Estadual de Montes Claros, Montes Claros, Brazil in 2010 and the M.S. degree in computer science from Universidade Federal de Pernambuco, Recife, Brazil, in 2015. Currently, he is a Ph.D. student at Universidade Federal de Pernambuco.

He is IT Analyst at Instituto Federal do Norte de Minas Gerais since 2010. His research interest includes machine learning, educational assessment, domain-driven data mining and models interpretability.

Paulo J. L. Adeodato received the B.S. degree in electronics engineering from Universidade Federal de Pernambuco, Recife, Brazil, in 1982, the M.S. degree in computer science from Universidade Federal de Pernambuco, Recife, Brazil, in 1991 and the Ph.D. degree in mathematics from King’s College London, London, United Kingdom, in 1997.

From 1982 to 1988, he was Production Manager with Digirede Computer Manufacturing. From 1999 to 2021, he was co-Founder, Partner and Chief Innovation Officer with Neurotech. From 1996, he is a Member of the Board of Associates of CESAR (Center for Advanced Studies and Systems of Recife). Since 1997, he holds a tenure-track position as professor with Centro de Informática da Universidade Federal de Pernambuco, Recife, Brazil. His research interest includes the development decision support systems with explainable artificial intelligence, quality assessment of binary classifiers, and startup entrepreneurship and investment. In 2017, he spent 3 months as Visiting Scholar with Stanford Graduate School of Education, Stanford, CA, USA. In 2018, he was a Visiting Professor in Stanford School of Medicine, Stanford, CA, USA, for 12 months.

Prof. Adeodato’s team awards include the 1st Place Winner of the FINEP Innovation Award for Small Entrepreneurship in the Brazilian Northeast (2013, Brazilian Innovation Agency), 1st Place in the Time Series Forecasting Competition (IEEE + Kaggle, 2012, Madrid, Spain), the 2nd Place in the NN3 Time Series Forecasting Competition (IEEE - IJCNN 2007, Orlando, USA), and the 2nd place in the PAKDD Data Mining Competition (2007, Nanjing, China).