Aalto University

OPEN ACCESS

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Rossi, Sippo; Rossi, Matti; Upreti, Bikesh Raj; Liu, Yong

Detecting political bots on Twitter during the 2019 Finnish parliamentary election

Published in:
Proceedings of the 53rd Annual Hawaii International Conference on System Sciences, HICSS 2020

Published: 01/01/2020

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY-NC-ND

Please cite the original version:
Rossi, S., Rossi, M., Upreti, B. R., & Liu, Y. (2020). Detecting political bots on Twitter during the 2019 Finnish parliamentary election. In T. X. Bui (Ed.), *Proceedings of the 53rd Annual Hawaii International Conference on System Sciences, HICSS 2020* (pp. 2430-2439). (Proceedings of the Annual Hawaii International Conference on System Sciences; Vol. 2020-January). Hawaii International Conference on System Sciences.
http://hdl.handle.net/10125/64040

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Detecting Political Bots on Twitter during the 2019 Finnish Parliamentary Election

Sippo Rossi
Aalto University
sippo.rossi@aalto.fi

Matti Rossi
Aalto University
matti.rossi@aalto.fi

Bikesh Raj Upreti
Aalto University
bikesh.upreti@aalto.fi

Yong Liu
Aalto University
yong.liu@aalto.fi

Abstract

In recent years, the impact of bots used for manipulating public opinion has become an increasingly prevalent topic in politics. Numerous sources have reported about the presence of political bots in social media sites such as Twitter. Compared to other countries, the influence of bots in Finnish politics has received little attention from media and researchers. This study aims to investigate the influence of bots on Finnish political Twitter, based on a dataset consisting of the accounts following major Finnish politicians before the Finnish parliamentary election of 2019. To identify the bots, we extend the existing models with the use of user-level metadata and state-of-art classification models. The results support our model as a suitable instrument for detecting Twitter bots. We found that, albeit there is a huge amount of bot accounts following major Finnish politicians, it is unlikely resulting from foreign entities' attempts to influence the Finnish parliamentary election.

1. Introduction

Nowadays, many organizations and individuals attempt to influence people by spreading propaganda in social media through large networks of bot accounts [1, 2]. There are multiple examples of bots being used to distort political discussions on Twitter. One of the most notable cases is the 2016 US presidential election, where an organization linked to the Russian government has been accused of striving to manipulate the elections by spreading fake news or biased content via Twitter bot accounts [2, 3]. In this light, a number of studies have delved into the detection of bot accounts through developing and testing new bot detection methods. Based on synthesizing key factors for bot detection reported in previous studies, the study developed an integrated framework for bot detection.

Specifically, this study aims to demonstrate how bots that are being used to influence politics on Twitter can be identified using machine learning approaches. To demonstrate the application of the method, we identified the bots that existed before the Finnish parliamentary election in April 2019 using user-level metadata. Noticeably, recent publications have found evidence of bots being used to influence opinions in countries such as the United States [3], Japan [4], Brazil [5] and Russia [6]. Similar studies have not been conducted in Finland, albeit there is already evidence of at least one large but inactive Finnish Twitter botnet according to a researcher at F-Secure [7, 8]. In other words, our study seeks to answer the following two research questions, including:

RQ1: What are the important features that can be used to identify bots?

RQ2: Do the bots have an impact on Finnish politics?

To answer the research questions, we first develop a model that can predict bots using machine learning methods. Once the bots are identified, we assessed the impact in terms of visibility and popularity of politicians followed by these bots.

This paper contributes to the growing information systems science and political data science literature on the use of bots and information systems to influence voters. The study also adds to bot detection literature by evaluating the feasibility of using a limited set of profile metadata features in a supervised machine learning bot detection model. As a part of the research project to detect bot's effect on ongoing European elections, we deem the study addresses a timely and important topic, as there is evidence of attempts to use bots to influence voters during recent European elections [9, 10].

2. Related research

We analyze the related research in three parts. The first part looks at how previous research has classified bots and provides a clear definition of key terms and concepts. The second part analyzes methods that have been used to detect bots in Twitter-related research and provides a background and benchmarks for the bot detection model proposed. The third and last part covers literature on the use of bots in political

URI: https://hdl.handle.net/10125/64040
978-0-9981331-3-3
(CC BY-NC-ND 4.0)

influencing during recent years to support the findings and assumptions made.

2.1. Terminology and the definition of a bot

A bot can be defined as an account that is operated fully or partially by a program. Thus, at least some parts of a bot account's activities are automated. Examples of these include bots belonging to like farms that are used on social media to increase the number of followers of an account or likes of a particular post. However, they are prone to detection and thus, deletion. More advanced bots adjust their content dynamically based on the behavior of other accounts, making them more difficult to detect even if the bot is still operated solely by a program. The most sophisticated bots are such that humans control parts of their activities, such as content creation, which blurs the line between the bot and a human user. When properly operated, these hybrid bots are almost invisible to automatic detection mechanisms, according to Grimme et al. [11]. Some bot accounts are inactive, also known as sleeper bots [12]. The accounts are 'quiet' most of the time before being activated e.g. to spread spam.

On Twitter, bots can be divided into benign and malicious bots [13]. The benign bots adhere to Twitter's rules and guidelines and are clearly distinguishable from human accounts usually by name or description. Conversely, malicious bots participate in activities that are not permitted by Twitter and rarely disclose the fact that they are operated by a program. Typical use cases include artificially boosting the number of followers, likes or retweets and directing or blurring discussions as well as spreading spam or content that supports a certain cause. Both types include bots ranging from simple content sharing accounts to human-like *social bots* that participate in discussions and create original content. The phrase of social bot here refers to a bot that is meant to mimic human behavior [12] by communicating and interacting with human users [14].

An important subtype of Twitter bots is political bots, which are specifically designed to participate in political discourse or to promote a certain ideology, organization, or individual [12]. In most cases, a political bot will have no references to it being a bot but may attempt to mimic human behavior in order to avoid detection and to influence other users.

2.2. Detecting bots on Twitter

2.2.1. Simple versus complex models. As algorithms that control bots become more advanced, so do the bot detection algorithms. In literature, bot detection models range from the very simple ones that are based on analyzing one piece of metadata to those that use ensemble methods to analyze large feature sets including a mix of metadata, tweeting behavior, and content data.

Past studies on bot detection have been to some extent restricted to bots with a specific feature. For instance, Beskow and Carley [15] managed to identify specific automatically generated bot accounts based on a single piece of metadata, the profile name, with approximately 95%-99% accuracy depending on the algorithm used. However, this type of approach results in a very narrow use and the aforementioned model could only detect bot accounts that have an account name consisting of a randomly generated string of 15 characters and more than likely to miss out the bots with different characteristics. However, as Beskow and Carley [15] propose, a tool-box approach where multiple different models are combined can make even the simple models an important contribution to more advanced bot detection models.

A number of bot detection models looked into various characteristics of accounts by combining metadata and behavior features to identify bots (e.g. [16, 17]). One notable issue hinders the reusability of these models. Many of these models rely on some form of natural language processing, sentiment analysis techniques [14] or a specific list of keywords [6, 16, 18]. This restricts their applicability to a particular language and region as well as an event such as an election, due to certain themes and hashtags being important only in that specific context. Bots have been evolving rapidly during the past few years to a point that they may be difficult even for a human to distinguish them from real users [19]. There is a need to update bot detection algorithms, since a workable algorithm today may prove to be ineffective after a couple of years.

2.2.2. Feature space selection. Machine learning methods represents the key approach used in early bot detection literature, in which an essential aspect of work is to determine the optimal feature space that boost bot prediction performances. There are two main considerations in the selection of bot detection features. Firstly, the features should be added only if they improve the accuracy of bot detection. Secondly, the features must not make the data collection phase overly time-consuming, since Twitter's API has strict rate limits.

In previous studies, the most common classes of features used in bot detection include metadata-based features and tweeting characteristics-based features [6, 8, 18]. User profile provides a large amount of metadata while tweets offer useful information, albeit

being limited to a certain number of characters (280). Based on these features, the amount of analyses that can be performed is vast. Other classes of features, such as keywords, are not included in the analysis, because these features restricted the applicability of the model to a specific event.

Metadata-based features can be divided into two different branches. Intuitively, metadata extracted from a profile gives information on the account, while metadata from tweets gives a combination of information from the profile posting it as well as the tweet itself [20].

Metadata that can be extracted from Twitter include basic profile information such as name, description, and number of friends. An examination on whether different pieces of profile information are blank or at default contributes to a collection of binary features, such as a variable of whether or not the profile picture has been added [6, 16]. The more fields are left at default, the more likely the account is a bot [6]. Data on the number of users that the profile is following, the number of followers and ratios of these are also often used in prior bot detection studies [6, 8, 18, 20]. Profiles that have none or a few followers, but follow many profiles are suspect [8]. Lastly, the contents of the textual metadata can be analyzed and used to classify bots for instance by inspecting the length or frequency of certain keywords in the description or name [8, 15, 18].

Earlier findings suggest that a combination of both metadata and content features yields optimal results [3, 18]. Hundreds of different features can be derived from Twitter's metadata and content data, making it a matter of preference on which ones to choose. Examples include counting the number of hashtags, URLs, and instances of specified keywords in the name or description of an account.

The model proposed in this study utilizes metadata-based features only and therefore, they are examined more thoroughly than content-based and other types of features. Further, unlike tweet content-based features, metadata-based features are more generalizable across different linguistic context. Table 1 illustrates some of the features that have been used in previous papers [6, 17, 18]. Unsurprisingly, the most common features are the ones that are directly related to how Twitter functions, default profile values, with the number of followers, friends, tweets, and retweets being examples of these.

Table 1. Summary of key features for bot detection used in prior literature

Binary features	Profile information features	Ratio features	Metadata content features
Defaults: - Profile image - Background image - No user description **Other:** - Profile verified - Location specified - No friends - No tweets	**General:** - Number of followers - Number of friends - Number of tweets - Number of likes - Age of account - Account language **Length:** - Profile name - Profile description	**Activity:** - Ratio of following and followers (FE/FI) - Reputation (FE/(FI + FE)) - Given likes per friend - Given likes per follower **Account age:** -Friends/Account age - Following rate (FI/AU)	**Bot check:** - Name contains bot - Description contains bot **Other content:** - Number of # in description - Keywords in description - URL(s) in description

In Table 1, the features are grouped into four types. Some of the most commonly used features are found in the first group as binary features. Based on the popularity, it can be assumed that they are appropriate for bot accounts detection despite their simplicity. Binary features are designed to check whether profile customization options, such as the profile image and background image, are left at default [6, 17, 18].

The second group also contains many of the prevalent features in bot detection models. These features are often numerical variables, many of which are related to how popular a Twitter account is and how actively it is used. Particularly, the numbers of followers, friends, tweets, retweets, and likes were often investigated [6, 17, 18]. Another commonly used feature is the length of the description text, which cannot be obtained directly from Twitter but can be calculated easily from the metadata [6, 17, 20].

The third group of features is ratios that can be obtained from the same metadata. When compared to the two previous groups, the ratio features offer more variety as they are not based on Twitter's built-in attributes. Followers-to-friends ratio is a common ratio feature used in many previous studies [6, 18, 20]. In the model created by Fernquist et al. [18], the top features for bot detection include multiple of ratios, with examples being given likes per friend, followers-friends ratio, and number of likes per followers.

The last group consists of the features deriving from the contents of different attributes. Features in this group are occasionally used in earlier studies. Two of the features in this list simply check whether an

account is a bot according to the profile description or name by looking if the fields contain the word "bot" [15]. The rest of the features relate to an examination of URLs, hashtags ,or other keywords [20].

Because ratio features were among the best performing features widely used in early studies [6, 18, 20], we include several of them in the study alike.

2.2.3. Classification methods. Because Twitter, like most of the social media sites, actively attempts to detect and disable bot accounts, the creators of bots have responded by making bots behave more like humans. Consequently, the selection of features as well as preparing the training data has become more demanding and for a model to stay up to date, feature engineering and adding new training datasets is needed [20].

Both supervised [6, 15] and unsupervised [16, 21] machine learning models have been used in bot detection research. The drawback of supervised learning is that creating a labeled dataset for training the model either requires a large amount of manual labeling [6] or using a pre-labeled dataset, which may limit the applicability of the model as the datasets most likely represent only a fraction of the possible behavior of bot accounts in Twitter. Unsupervised learning models can detect novel bot behavior that may get past a supervised model [16], as the supervised models can only detect bots that are similar enough to the dataset that was used to train it. However, the results of unsupervised models are more difficult to validate due to the absence of labeled data.

Past studies indicated that supervised models are better suited for analyzing topical datasets that are collected from Twitter's streaming API [6]. Twitter's API allows performing searches and collecting the data on tweets that contain for certain keywords or hashtags, which is particularly useful when analyzing political discourse that is related to a specific topic, such as an election [18, 22]. Since campaigns, political parties, candidates and users use hashtags to make their tweets visible when commenting on specific topics, it is more efficient to mine data on a topical level with the keyword search instead of first collecting a large dataset of Twitter accounts and then analyzing the content of their tweets.

2.3. Use of bots in political influencing

Previous studies illustrated that Twitter-based computational propaganda has been used by organizations and governments across the world [12]. There are several hypothesized goals of the creators of bots. These range from increasing the partisanship of a population or advancing a cause that the creator of the bots supports. [23] noted that "it is an effective non-military means for achieving political and strategic goals." Measuring the successfulness of political bots is difficult as it is hard to quantify the impact that they have had for example, on voting behavior [23]. Nevertheless, the prevalence of computational propaganda campaigns would suggest that they are viewed as a functional tool that does have an effect on the target audience [23].

More measurable and easily achievable targets include manipulating the popularity and visibility of tweets by liking, following, and retweeting content with a botnet. These methods can cause a particular hashtag to trend thus, pushing it higher into the feeds of other Twitter users. Other goals may be to make an opinion seem more popular than it actually is or to bury actual discussions or factual information by making it difficult to follow. Concrete examples include spamming pro-government tweets or flooding search results related to protests with meaningless content making it more difficult for human users to find and participate in discussions [24].

Two earlier studies monitored bot activity in Germany during a state parliament as well as Federal presidential election [10] and federal election during 2017 [9]. Bots represented around 7 - 11% of the accounts and bot-driven content represented 7.4 - 9% of all traffic during the German elections [9, 18]. These are modest numbers and in line with Twitter's estimate of bots accounting for approximately 10% of Twitter activities. The main reason for concern is that the bot activity was skewed towards supporting the alt-right movement and was possibly produced by accounts outside of Germany [9]. As an extreme example, in Russia, Stukal et al, [6] reported that up to 85% of the daily tweets containing political keywords were posted by bot accounts during 2014-2015. Obviously, there are regional differences in the prevalence of bot accounts [12].

Finland, the focus of this research, may also be affected by bot account, considering i) the current growth of Euroscepticism, ii) rise of right-wing political movements alongside Finland's historical relations and iii) proximity to Russia that may provide a more fertile foundation for bot activity than for example Sweden.

3. Methodology

This study employed an unorthodox approach to collect Twitter data as the dataset is compiled from individual accounts' followers, which differs from the more traditional methods by collecting all tweets (and associated account metadata) that use specific hashtags or keywords are gathered through Twitter's Streaming

API. This method is appropriate since the proposed bot detection model only requires metadata. The benefits of this approach include that it allows detecting both dormant bots as well as those that do not use specific hashtags or words that the streaming method queries for.

The primary tools used in data collection and formatting phase were the statistical programming language R and its "rtweet" package, which is an "R client for accessing Twitter's REST and stream APIs".

The study analyzed the Twitter accounts of several politicians and their followers on Twitter. The profiles were selected based on several heuristically chosen criteria to ensure that as many political parties as possible were represented and that a sufficient amount of data was collected. At least a member of parliament was taken from each of the current coalition parties as well as from all parties that have support of over 5%, albeit a maximum of two per party. Furthermore, only accounts with over five thousand followers were picked. Lastly, some prominent politicians with over 10 thousand followers were selected even if they do not match the other criteria as several influential political figures would otherwise be excluded. Table 2 shows the selected politicians.

Table 2. Summary of selected politicians

Name	Political party	Username	Followers (K)*
Alexander Stubb	National Coalition Party	@alexstubb	370
Sauli Niinistö	National Coalition Party	@niinisto	159
Juha Sipilä	Centre Party	@juhasipila	126
Anne Berner	Centre Party	@AnneBerner	21.7
Pekka Haavisto	Green League	@Haavisto	130
Ville Niinistö	Green League	@VilleNiinisto	84.3
Paavo Arhinmäki	Left Alliance	@paavoarhinmaki	109
Li Andersson	Left Alliance	@liandersson	76.5
Antti Rinne	Social Democratic Party	@AnttiRinnepj	25.6
Sanna Marin	Social Democratic Party	@MarinSanna	14.3
Jussi Halla-aho	Finns Party	@Halla_aho	14.5
Laura Huhtasaari	Finns Party	@LauraHuhtasaari	13.7
Sampo Terho	Blue Reform	@SampoTerho	7.6
Paavo Väyrynen	Seven Star Movement	@kokokansanpaavo	10

*Number of followers at March 2019

The sample consists of 14 politicians from 8 different parties ranging from liberal to conservative and left-wing to right-wing, including the current president and three ministers as well as 6 party leaders. Many small parties were left out by this approach, of which respective politicians did not have accounts or had much fewer followers. As a result, over 1.1 million Twitter accounts were collected as followers of these politicians, but the number was reduced to approximately 550,000 after filtering out duplicate accounts. The duplicates were a result of the fact that many Twitter users were following multiple selected politicians.

The binary features were calculated from corresponding attributes where a setting left at default or blank equals 1 and a nondefault 0. The ratio features were created similarly by calculating the values from the profile information metadata and then placed into new columns. Ratio calculations that resulted in NaN (not a number) or Inf (infinite), were replaced with a zero.

3.1. Creating training data

Based on the findings of previous research and datasets available for training the model, a selection of 11 features was picked for testing the first version of the model. The feature space consists of four binary features, four profile information features, and three ratio features.

Initially, the model was trained with the cresci-2017 dataset [25], which contains over 13,000 labeled accounts divided into groups of social spambots, traditional spambots, fake followers, and genuine accounts. The training dataset was balanced to include 3,000 randomly sampled bot accounts and 3,000 randomly sampled genuine accounts from the cresci-2017 dataset.

To find a suitable algorithm for the prediction, the Linear Discriminant Analysis (LDA), Classification and Regression Tree (CART), K-nearest neighbors (KNN), Support-vector machine (SVM) and Random Forest algorithms were tested. Out of these Random Forest performed the best, although there were signs of either the training data not representing the variety of real data or that the model being overfitted as the accuracy was over 97% or 98% on most runs. This issue was ignored, as the model was deemed sufficiently accurate for the first phase where the goal was mainly to speed up the training data creation by

manually validating list of potential bot accounts from the prediction results. The model was then tested on a sample of 5,000 accounts from the dataset that was collected for this study.

After manually inspecting on Twitter the accounts that the model labeled as bots, it was evident that the model had difficulties distinguishing bots and genuine accounts. Particularly, accounts that were apparently created by people trying out Twitter without becoming active users were prone to be labeled as bots due to the similarity in the account behavior. In most cases, the easily distinguishable bots were following approximately 20-100 accounts, had 0-2 followers and little to no tweets, retweets or likes.

Based on the performance of the first version of the model, it was apparent that the cresci-2017 dataset was unsuitable for training a model that could accurately distinguish bots from humans based on metadata. One possible explanation for such performance is the fact that the training data used in the model had only very clear examples of bots and genuine accounts, where the behavior in terms of tweets, retweets, likes and ratios of followers and following differed widely depending on whether the account was a bot or not. However, this does not reflect the actual behavior of accounts where in some cases even with quantitative and qualitative assessment it is difficult to label an account accurately as either a bot or a human.

By manually labeling a set of accounts from the dataset consisting of followers of the Finnish politicians, a new training dataset that represents the actual distribution and behavior of the accounts of the target dataset was created. The training data was created by checking and verifying the accuracy of 2,000 accounts predicted to be bots by the first model. The results were that out of these accounts 1,336 were accurately labeled as bots, as they were either bots or accounts exhibiting extremely bot-like behavior while 664 were actually humans or accounts that were impossible to determine as belonging to either group.

A qualitative approach was employed for classifying the accounts as either bots or humans. The classification started by inspecting the profile information of the account. Common signs of a bot were the name or description of the account, which often included Russian or Arabic and or a seemingly random string of characters and numbers coupled with the account following 21 other Twitter users, which is the default number of recommended users to follow given by Twitter when creating a new account. Other possible predictors included in this step are the profile image and banner as well as the age of the account. As a second step, the tweets and retweets were checked when available to see what kind of activity the account has and what other accounts it interacts with. As the third step, the accounts that the possible bot was following were inspected to find discrepancies. For example, a user following mainly seemingly random foreign accounts coupled with one Finnish politician or if it was following exactly 21 very popular Finnish accounts were usually the best predictors of an accurate classification as a bot even though the machine learning model did not look for these. If after the three first steps the account was still too ambiguous for classification, the likes and followers were checked for bot-like behavior.

During this process, several interesting findings were made, which can be used later in the analysis of the whole dataset. Firstly, most of the bot accounts were dormant as well as possibly a part of a follower boosting operation. Secondly, most of the bots were difficult to label as political bots as it is not sure whether they were created to boost the followers of a particular politician or if it followed them by coincidence based on Twitter's recommendations. Commonly, shared characteristics among bots included that they barely engaged with content or interacted with other users and that they followed a random group of 21 accounts, which most likely are those suggested by Twitter during the creation of the account [7, 8]. Peculiar accounts that they often followed included less well-known US politicians, an obscure game called Growtopia and a niche Finnish newspaper called Markkinointi and Mainonta.

3.2. Building the bot prediction model

The second version of the bot detection model differed from the previous one mainly in how the splitting of the training and validation data was done, what parameters and algorithms were used as well as how many features were included.

New features could be added to the second version of the model as the training data was no longer a limiting factor. By including the age of the account, and two ratio features derived from comparing the profile information to the age of the account, the number of features was increased from 11 to 14.

Several variants of the Random Forest algorithm were tested, but the standard version still performed optimally and was selected for the final model. The model was trained with a randomly sampled set of 500 bots and 500 humans from the new manually labeled dataset. The remaining 1000, with 836 bots and 164 humans, were used in the validation of the

performance. The final version of the bot detection model has an accuracy of 83% with only slight changes after multiple runs and small variations in parameter settings. Table 3 lists the most important statistics for assessing the performance.

Table 3: Performance of the bot detection model

Metric	Value
Accuracy	0.837
Recall	0.846
Specificity	0.793

Table 4: Features ranked

Rank	Feature	Importance
1.	Following	100.00
2.	Following to age of account	61.00
3.	Age of account	56.57
4.	Followers to following	22.87
5.	Likes to following	15.68
6.	Tweets	15.18
7.	Likes to age of account	11.95
8.	Followers	10.05
9.	Default profile image	7.11
10.	Likes	6.34
11.	No description	4.72
12.	No location	3.61
13.	No banner	2.56
14.	Likes to followers	0.00

In terms of feature importance, the top features were a mix of profile information and ratio features, while the binary features were all in the bottom half of the feature ranking. Table 4 contains the full ranking of the features. Based on this, the model gives much weight to the number of accounts that an account is following, since the two top features are related to the following attribute. This is somewhat problematic for our overall goal of political bot detection as it implies that the model is best at detecting dormant bots and bots belonging to follower farms. These accounts can be political bots, but in many cases determining if they are following politicians on purpose or by coincidence is difficult. This is because the popular politicians often appear on the top of the recommended accounts to follow in Finland.

4. Findings and discussion

4.1. The proposed bot detection model

The bot detection model proposed in this study demonstrated that metadata alone is sufficient for classifying at least spambots and bots that belong to follower farms. The primary benefit of a model based on metadata is that the data collection is much quicker as 90,000 accounts' information can be retrieved every 15 minutes. Therefore, a model that uses metadata works particularly well when studying countries that have a small population, since then even the most popular Twitter users are likely to have a manageable number of followers. In other words, due to the limited number of users in these countries, it is possible to gather comprehensive datasets for analysis in short periods. Furthermore, analyzing entire populations instead of samples is feasible with a purely metadata-based model, contrary to models that use tweet data, where the number of accounts to analyze is restricted by Twitter's streaming API's rate limits.

Regarding the selection of the feature space and algorithm, most of the results were in line with the reviewed literature, although some of the results were surprising. Random forest was the optimal classification algorithm, which was the result in several other models as well [18]. While ratio features had high feature importance as suggested by previous research, the binary features did not despite their popularity in earlier models. Overall, the performance of the model was below most of those listed in the literature review, but as stated earlier direct comparison is difficult due to the differences in the goals of the models.

4.2. Bots Counts in Finnish political Twitter

Based on our bot detection model, we predicted the total number of bots in the dataset consisting of the 558,983 followers of the 14 Finnish politicians was formatted to match the training dataset. The model predicted that out of the dataset approximately 36.6% are bots. Since the model's accuracy is 83%, out of the 204,426 accounts classified as bots it can be assumed that 169,673 should be the real number of bots when not taking into consideration the accounts labeled as humans that in reality, are bots. Therefore, the percentage of bots in reality is likely to be closer to 30% based on the results and the accuracy of the model.

4.3. Influence of Bots in Finnish political Twitter

Overall, the findings of the study do not support the notion that Finland and Finnish politics would be the target of internal or external bot influencing campaign, due to most of the bots having almost no activity besides following popular accounts. This finding is line with a recent announcement made by

Supo, the Finnish Security Intelligence Service, which stated that it has not found evidence of foreign entities attempting to influence the elections [26].

Despite few political bots, over 150,000 bot accounts following Finnish politicians on Twitter were identified. Although these bot accounts do not interact much with other accounts, they still help the politicians that they follow by two ways. Firstly, they artificially inflate the number of followers a politician has making them possibly more popular than they actually are. Secondly, they help increase the visibility of politicians, since being followed by many promotes an account over other less popular accounts in Twitter's "who to follow" suggestions. Consequently, bot accounts that were created for an entirely different purpose may unintentionally follow politicians when they follow their accounts based on Twitter's recommendations.

The primary impact that the bots have on Finnish political Twitter is related to increase the visibility and perceived popularity of the politicians' accounts. Considering a low utilization of Twitter as a medium for political debate in Finland, the possible effects the bots that may have had on voters may be negligible. Nevertheless, one metric for measuring a politician's popularity that can be used to predict election results is how many followers they have on different platforms and how much their audience engages with them [27]. Therefore, even if the impact on actual voting behavior is minimal, the presence of bots may manipulate perceptions, influence predictions and damage the validity of social media engagement as an indicator of actual popularity.

When inspecting the scores of individual politicians, Pekka Haavisto and Alexander Stubb had the highest percentages of bot followers, with both at above 30%, which is beyond Twitter's own estimates of 5-10% accounts being bots. The strong bot presence in Haavisto's Twitter follower base was subject to debate already in 2017 during his presidential election campaign [28]. Previous analysis attributed the bot followers to a result of a sudden increase in bots promoting the game Growtopia and Twitter's recommendations boosting Haavisto, which is similar to the findings of this study.

Alexander Stubb, the other notable example of a politician benefitting from the added visibility, has acquired the largest absolute number of bot followers. Many of the bots did not follow any other politicians besides Stubb, which is likely due to his strong presence in Twitter as the 3rd most followed account in Finland.

Contrary to findings elsewhere [4, 9], the candidates most likely to be linked to the Finnish alt-right movement Laura Huhtasaari and Jussi Halla-aho had the lowest percentage of bot followers. However, this is not surprising when taking into consideration that they also have the lowest number of followers from the sample of accounts inspected, which means that they do not attract bots that follow accounts by default based on Twitter's recommendations.

5. Conclusions

The goals of this study were to develop a new supervised machine learning bot detection model to investigate if Twitter bots were used to influence the 2019 Finnish parliamentary election and to test a new approach for Twitter bot detection. The developed model was used to estimate the number of bot followers that a sample of the most popular Finnish politicians have in their follower base.

The dataset used in the study consisted of 550,000 unique accounts out of which roughly 169,600 were classified as bots. The metadata-based model was found to be feasible for classifying bots on Twitter and the predictions of the model were used to assess if bots were utilized during the 2019 Finnish parliamentary election. The findings imply no evidence of attempts to influence the elections via Twitter bots. Although the bots increased the visibility of some politicians and made them seem more popular, the bots are unlikely to have had much effect due to their passive behavior.

This study holds important implications for both the researchers as well as practitioners. Our study explores a number of primary meta data-based features as well as ratio-based profile features to predict bots in Twitter. This approach provides better coverage of the profile characteristics, and it is generalizable to a wide variety of context due to the linguistics independence.

To the best of our knowledge, this study is the first in studying the presence and influence of bots in a Finnish context. Our results imply that the bots are surfacing in the Finnish domain. Even though we did not find the bots to have a significant impact, we cannot predict how this could change in the future. These results should be of interest not only to researchers, but also to politicians and users of social media in Finland.

Lastly, our results also highlight the influence that Twitter's suggestions can have on the number of followers that popular accounts have. These results indicate that profiles followed by bots are likely to attract more bots, further inflating their number of followers and perceived popularity.

5.1. Limitations

Like all studies, our study also has limitations. First, the approach used in the selection of politicians and data collection phase as well as the choice of features in the machine learning model introduced some constraints to the analyses that could be performed. Our sampling approach ignores the user accounts that do not follow politician yet remain politically active. Although it was possible to determine if an account is a bot based on metadata, the collected data did not enable examining the content that they interacted with or spread via tweets, retweets, and likes. However, it is worth noting that most of the bots detected are not actively creating or distributing content. Lastly, politicians with much higher or lower percentages of bot followers may have been omitted from the sample.

5.2. Suggestions for further research

To further understand the use of bots in the Twittersphere, the model could be reused during future elections by collecting new datasets. This would be particularly interesting due to the Finnish Security Intelligence Service's suggestion that the EU elections are likely to be a more attractive target for external influencing attempts than the Finnish parliamentary election [26].

To analyze the efficiency of Twitter's own bot detection and removal practices, the rate at which accounts labeled as bots are removed from the social media site can be followed. In addition, changes in the activity of the bots can be monitored by inspecting how the attributes such as a number of tweets and likes changes over time. Especially interesting would be to find evidence if some of the accounts were sleeper bots waiting for activation.

6. References

[1] D. W. Nickerson and T. Rogers, "Political campaigns and big data," *Journal of Economic Perspectives,* vol. 28, no. 2, pp. 51-74, 2014.

[2] Twitter. "Update on Twitter's review of the 2016 US election." https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html.

[3] A. Bessi and E. Ferrara, "Social bots distort the 2016 US Presidential election online discussion," 2016.

[4] F. Schäfer, S. Evert, and P. Heinrich, "Japan's 2014 General Election: Political Bots, Right-Wing Internet Activism, and Prime Minister Shinzō Abe's Hidden Nationalist Agenda," *Big data,* vol. 5, no. 4, pp. 294-309, 2017.

[5] C. A. D. L. Salge and E. Karahanna, "Protesting Corruption on Twitter: Is It a Bot or Is It a Person?," *Academy of Management Discoveries,* vol. 4, no. 1, pp. 32-49, 2018.

[6] D. Stukal, S. Sanovich, R. Bonneau, and J. A. Tucker, "Detecting bots on Russian political Twitter," *Big data,* vol. 5, no. 4, pp. 310-324, 2017.

[7] E. Gallagher. "Visualizations of the Finnish-themed Twitter botnet." https://medium.com/@erin_gallagher/visualizations-of-the-finnish-themed-twitter-botnet-bfc70c6f4576.

[8] A. Patel. "Someone Is Building A Finnish-Themed Twitter Botnet." https://labsblog.f-secure.com/2018/01/11/someone-is-building-a-finnish-themed-twitter-botnet/.

[9] F. Morstatter, Y. Shao, A. Galstyan, and S. Karunasekera, "From alt-right to alt-rechts: Twitter analysis of the 2017 german federal election," in *Companion of the The Web Conference 2018 on The Web Conference 2018*, 2018: International World Wide Web Conferences Steering Committee, pp. 621-628.

[10] L.-M. N. Neudert, "Computational propaganda in Germany: A cautionary tale," *Computational Propaganda Reserach Project, Paper,* vol. 7, p. 2017, 2017.

[11] C. Grimme, M. Preuss, L. Adam, and H. Trautmann, "Social bots: Human-like by means of human control?," *Big data,* vol. 5, no. 4, pp. 279-293, 2017.

[12] S. C. Woolley and P. N. Howard, "Computational propaganda worldwide: Executive summary," *Working Paper,* no. 11. Oxford, UK, p. ProjectonComputationalPropaganda, 2017.

[13] R. J. Oentaryo, A. Murdopo, P. K. Prasetyo, and E.-P. Lim, "On profiling bots in social media," in *International Conference on Social Informatics*, 2016: Springer, pp. 92-109.

[14] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, "Botornot: A system to evaluate social bots," in *Proceedings of the 25th International Conference Companion on World Wide*

Web, 2016: International World Wide Web Conferences Steering Committee, pp. 273-274.

[15] D. M. Beskow and K. M. Carley, "Its all in a name: detecting and labeling bots by their name," *Computational and Mathematical Organization Theory,* pp. 1-12, 2019.

[16] A. Minnich, N. Chavoshi, D. Koutra, and A. Mueen, "BotWalk: Efficient adaptive exploration of Twitter bot networks," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 2017: ACM, pp. 467-474.

[17] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," in *Eleventh international AAAI conference on web and social media*, 2017.

[18] J. Fernquist, L. Kaati, and R. Schroeder, "Political Bots and the Swedish General Election," in *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2018: IEEE, pp. 124-129.

[19] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Communications of the ACM,* vol. 59, no. 7, pp. 96-104, 2016.

[20] B. Wang, A. Zubiaga, M. Liakata, and R. Procter, "Making the most of tweet-inherent features for social spam detection on twitter," *arXiv preprint arXiv:1503.07405,* 2015.

[21] N. Chavoshi, H. Hamooni, and A. Mueen, "DeBot: Twitter Bot Detection via Warped Correlation," in *ICDM*, 2016, pp. 817-822.

[22] B. Kollanyi and P. N. Howard, "Junk news and bots during the German parliamentary election: What are German voters sharing over Twitter," ed: Oxford University: Comprop Data Memo, 2017.

[23] C. Bjola, "Propaganda in the digital age," ed: Taylor & Francis, 2017.

[24] P. Suárez-Serrato, M. E. Roberts, C. Davis, and F. Menczer, "On the influence of social bots in online protests," in *International Conference on Social Informatics*, 2016: Springer, pp. 269-278.

[25] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017: International World Wide Web Conferences Steering Committee, pp. 963-972.

[26] N. Simojoki. "“Ei se ulkopuolelta suunnatulta kampanjalta vaikuta” in Finnish." https://demokraatti.fi/ei-se-ulkopuolelta-suunnatulta-kampanjalta-vaikuta-asiantuntijat-eu-vaalit-houkuttelevampi-vaikutusyritysten-kohde/.

[27] J. DiGrazia, K. McKelvey, J. Bollen, and F. Rojas, "More tweets, more votes: Social media as a quantitative indicator of political behavior," *PloS one,* vol. 8, no. 11, p. e79449, 2013.

[28] F.p.s.b.c. (Yle). "Pekka Haavisto campaign concerned over suspected Twitter bots." https://yle.fi/uutiset/osasto/news/pekka_haavisto_campaign_concerned_over_suspected_twitter_bots/9988551.