Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Information Technologies

Bachelor Thesis

Identifying accounts that spread disinformation on Twitter

Adil Kussataiuly

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

BACHELOR THESIS ASSIGNMENT

Adil Kussataiuly

Informatics

Thesis title

Identifying accounts that spread disinformation on Twitter

Objectives of thesis

Main objective:
This research aims to identify networks of accounts that spread disinformation on Twitter.
Partial objectives:
1) To make an overview of the current state of the play in identifying Twitter accounts spreading disinformation and methods of detecting them.
2) To select a computational method for bots identification and prepare a dataset of tweets concerning a selected topic.
3) To run an experiment to detect accounts spreading disinformation
4) To interpret findings and formulate a conclusion.

Methodology

The methodology of solving the theoretical part of the diploma thesis will be based on the study and analysis of professional information sources. Based on the knowledge gained in the theoretical part of the work, the practical part will identify the spreading of propagandistic posts on the Twitter network by specific users and bots. Furthermore, experimental measurements will be performed using appropriate tools. The obtained data will be evaluated. Based on the synthesis of theoretical knowledge and the results of the practical part, the conclusions of the work will be formulated.

The proposed extent of the thesis

50 pages

Keywords

Twitter, fake, COVID19, accounts, bots, spread

Recommended information sources

BELLOVARY, Andrea; YOUNG, Nathaniel A.; GOLDENBERG, Amit. Left-and Right-Leaning News Organizations' Negative Tweets are More Likely to be Shared. 2021.

CALDARELLI, Guido, et al. The role of bot squads in the political propaganda on Twitter. Communications Physics, 2020, 3.1: 1-15.

GRUZD, Anatoliy; MAI, Philip. Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. Big Data & Society, 2020, 7.2: 2053951720938405.

GUARINO, Stefano, et al. Characterizing networks of propaganda on Twitter: a case study. Applied Network Science, 2020, 5.1: 1-22.

MOZUR, Paul, LEE MYERS, Steven, KAO, Jeff and THE NEW YORK TIMES. How bots and fake accounts push China's vision of Winter Olympic Wonderland. How Bots and Fake Accounts Push China's Vision of Winter Olympic Wonderland [online]. 18 February 2022. [Accessed 5 May 2022]. Available from: https://www.propublica.org/article/how-bots-and-fake-accounts-push-chinas-vision-of-winter-olympic-wonderland

WANG, Di; LU, Jiahui. How news agencies' Twitter posts on COVID-19 vaccines attract audiences' Twitter engagement: A content analysis. International Journal of Environmental Research and Public Health, 2022, 19.5: 2716.

Expected date of thesis defence

2022/23 SS – FEM

The Bachelor Thesis Supervisor

Ing. Miloš Ulman, Ph.D.

Supervising department

Department of Information Technologies

Electronic approval: 14. 7. 2022	Electronic approval: 27. 10. 2022
doc. Ing. Jiří Vaněk, Ph.D.	**doc. Ing. Tomáš Šubrt, Ph.D.**
Head of department	Dean

Prague on 13. 03. 2023

Declaration

I declare that I have worked on my bachelor thesis titled "Identifying accounts that spread disinformation on Twitter" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the bachelor thesis, I declare that the thesis does not break any copyrights.

In Prague on date of submission ______________________________

Acknowledgement

I would like to thank Ing. Miloš Ulman, Ph.D. for his advice and support during my work on this thesis.

Identifying accounts that spread disinformation on Twitter

Abstract

The spread of disinformation on social media platforms, particularly Twitter, has become a growing concern in recent years. This study aims to identify networks of Twitter accounts that spread disinformation by analysing a dataset of tweets related to specific global actions, such as elections, COVID-19, and war. Over 10,000 users were randomly selected from the dataset, and the Botometer® algorithm was used to identify potential bots. Various methods were used to detect Twitter accounts spreading disinformation, including network analysis, natural language processing, and machine learning algorithms. The Botometer® algorithm was selected as the computational method for bot identification in this study. A unique dataset was created by searching for particular keywords related to each topic of interest. The findings suggest that the number of tweets is a valuable variable in distinguishing low-activity bots from highly active users in the network. The results also showed that the group of users who tweeted about the election had the highest number of tweets but the lowest spammer value, while the groups of users who tweeted about COVID-19 and war had significantly fewer posts. This study confirms the relationship between bot score and the number of fake subscribers/spammers, providing valuable insights for detecting and preventing fraudulent behaviour on various online platforms. The practical implications of this research include improving the accuracy of bot detection systems and developing better tools for detecting and preventing disinformation on social media platforms. Ongoing research efforts in this area provide promising methods for mitigating the spread of disinformation and improving the accuracy and reliability of online data.

Keywords: Twitter, fake, COVID-19, accounts, bots, spread, elections, propaganda

Table of content

1 Introduction

In recent years, the spread of disinformation on social media platforms, particularly Twitter, has become a significant concern. Disinformation, or false and misleading information, can have serious consequences on individuals, groups, and even entire societies. It can manipulate public opinion, influence elections, and exacerbate social and political divisions. As such, identifying and detecting Twitter accounts that spread disinformation has become a crucial area of research for both academics and practitioners.

To tackle this issue, researchers and experts have developed various methods for identifying and detecting Twitter accounts that spread disinformation. One of the most common methods is network analysis, which involves identifying patterns and relationships between accounts that share disinformation. Researchers also use natural language processing techniques to analyze the content of tweets and identify patterns that suggest disinformation. Machine learning algorithms are also employed to analyze large volumes of Twitter data and identify patterns that suggest the presence of disinformation. Despite the ongoing research efforts, the task of identifying Twitter accounts spreading disinformation remains challenging, and further research is needed to develop more effective methods and tools.

Therefore, the goal of this study is to identify networks of accounts that spread disinformation on Twitter and investigate the characteristics and behaviour of these accounts. In order to achieve it, a dataset using keywords and hashtags related to specific events, including elections, COVID-19, and war was collected. A computational method for identifying bots, the Botometer® – a supervised machine learning classifier that uses attributes to discriminate between bot-like and human-like accounts – was used to detect accounts with bot-like behaviour. The findings of this study could have practical implications for individuals and organizations looking to mitigate the impact of fake accounts and improve the accuracy and reliability of their data on Twitter and other social media platforms.

2 Objectives and Methodology

2.1 Objectives

Main objective:

The aim of this research is to identify networks of accounts that spread disinformation on Twitter.

Partial objectives:

1) to make an overview of the current state of the play in identifying Twitter accounts spreading disinformation and methods of detecting them.
2) to select a computational method for bots identification and prepare a dataset of tweets concerning a selected topic.
3) to run an experiment to detect accounts spreading disinformation.

2.2 Methodology

The methodology of solving the theoretical part of the diploma thesis will be based on the study and analysis of professional information sources. Based on the knowledge gained in the theoretical part of the work, the practical part will identify the spreading of propagandic posts in Twitter network by specific users and/or AI bots. Furthermore, experimental measurements will be performed using appropriate tools. The obtained data will be evaluated. Based on the synthesis of theoretical knowledge and the results of the practical part, the conclusions of the work will be formulated.

3 Literature Review

3.1 Information

> "Information is information, not matter or energy."
> — Norbert Wiener, mathematician, and philosopher.

Regardless of the manner in which it is providing, information is a set of data. There are parallels to the notion, that originally among our ancestors, information was expressed through body language and verbal communication before evolving to the graphical cave paintings. Naturally, developing sophisticated information transfer methods required a long time. An inherent feature of typical human evolution is people's need to learn new things and interact with one another in some way. It became vital to create ways to transmit knowledge over distance as mankind advanced and its interactions with other advanced civilizations increased.

History states that writing first originated in Mesopotamia around the middle of the 4th millennium BC. Since that time, individuals have established the primary informational conduit for global communication. Writing has a huge cultural significance for humans. The earliest civilizations came into existence as a result of a confluence of elements, including geographic, social, and economic reasons as well as the development of writing. Therefore, it was difficult to imagine a future without writing. Reading and writing skills are the foundation of every society, and knowledge does not decay with time. The development of writing made it possible for information to be accumulated and reliably passed down to future generations. It was made feasible to gather and spread information with the creation of written text. Writing-educated civilizations advanced more quickly and to a higher degree in terms of both culture and economy.

Since 1983, when the World Wide Web first appeared, all of humanity's information flow has converged, creating enormous potential in all aspects of our existence. The quantity of information available to people on the Internet at this moment is nearly infinite. Although part of this material is accurate, some of it may even be phony, which is the unfortunate side of the technology advance. False news became simpler to disseminate with the introduction of the Internet, and as a result, this tendency started to cause severe issues in society more frequently. For instance, a news report purportedly published by The Associated Press that claimed Microsoft was purchasing the Catholic Church (The Times, 1994) became viral online in December 1994 and was mostly spread through email. It is said to have been the first online hoax to be seen by a large number of people.

On several occasions, the terms "fake news" and "fake media" have been used to characterize stories that the complaint disagrees with. People started using this keyword actively in searches in the second half of 2016, according to the Google Trends map. This year, social networks like Facebook's news feed and Twitter's microblogging platform brought the false news epidemic back into the public eye. As reported by the Pew Research Center (Gottfried et al. 2016), a considerable number of Americans use Facebook or Twitter to monitor news, and when paired with rising societal political polarization, as well as a phenomenon known as the "filter bubble" and a tendency to read mostly headlines – false news appeared to have an influence on the 2016 presidential election (The Guardian 2016).

As a result of the bogus news being more entertaining or more in line with expectations, it received more shares on Facebook than legitimate ones. The dissemination of false information on Facebook has increased threefold since 2016, referring to the research released in October 2020 by Digital New Deal (Kornbluh et al. 2020). Besides that, false news is frequently spread through websites that post news stories designed to catch reader’s interest and impersonate credible news organizations in order to gain their trust. Some of the "news" created on these websites occasionally makes its way to more reputable outlets.

> “If a lie is telling you something you want to hear, you’re more likely to think it’s true,”
> — Sharon Kaye, philosophy professor.

3.1.1 Social Media as a modern information channel

Formally speaking, social media are websites and software applications that enable users to share information and connect with other users via a smartphone or computer, according to the Cambridge Dictionary. If we separate the term social media, social (from the Latin word 'socius') states friend, while media (from the Latin word 'media') denotes methods of communication. To put it simply, social media may be considered of as a free-form space for public discussions.

Social media is widely used and significant, as evidenced by itself that majority of us are an active user of Facebook, WhatsApp, Twitter, YouTube, Instagram, and any other platforms. It is becoming a common occurrence in our daily lives. Let's first discover the features or qualities that make social media a special, wildly popular channel with broad use before we can comprehend the role of social networks in our lifestyle.

The ease of access social media gives its users is the key factor in social media's great success. Anyone may easily sign up to join the service if they have just any digital device with an internet connection. Social media has become even more prominent as a result of the Indian population's easy and affordable access to an internet service. A simple tap will connect you to anyone.

Text, music, and graphic elements may all be used by user to generate content. Whoever may launch personal blogs, audio podcasts, group conversations, private channels, and also exclusive group chats. Consequently, we have access to information in many categories and covering a wide range of subjects. Social networking promotes originality and innovation.

The fast nature of social media contributes to its extraordinary appeal and extensive use. Any post is seen by everyone in user’s friends list simultaneously in a short period of time. In fact, people may choose to share what someone has already shared. This will aid in becoming well-known or spreading like wildfire. Posts have frequently gained worldwide attention in latest days barely minutes after they have been uploaded, around several cases.

Everyone who uses a social media network can participate. In contrast to conventional media, such as newspapers, where material is only communicated, this one is more interactive. The functionality of social networking sites is reciprocal. On a real-time basis, you may interchange relevant data, discuss your experience, and offer your ideas and comments.

People are highly involved; therefore, the tools and information are always being modified and enhanced. The system is always adaptable to new needs and modifications. Users have sought a timely update, according to the software's perspective. This optimizes the system while eliminating the issues.

It's also important to remember that using social media is inexpensive. To access the social media realm, the user merely needs to pay for internet costs. Additionally, there are no fees associated with maintaining a presence on these networks. Besides this, you may promote and interact without spending a lot of money. Comparatively speaking, it is far more affordable than other forms of advertising or routes for interaction.

3.1.2 Definition of propaganda

Propaganda is a kind of conversation that aims to spread a specific agenda or perspective. It has the potential to persuade individuals or to regulate their attitude. False news or misinformation are frequently used in propaganda, and they may be really powerful in influencing public perceptions (Council of Europe).

Propaganda is commonly applied in politics to convince individuals to choose a certain political figure or idea. Traditional propaganda usually originates from a single source, like a company or business, and employs methodological approaches, which is drastically opposite to what is happening now. The following are a few examples of widely known propaganda tactics (Cuncic 2022):

- Manipulating information: entails altering or misrepresenting information in order to sway viewpoints. A political campaign, for example, may create false accusations around an opposition with the purpose of making them appear worse.
- False statistics: are a prevalent rhetorical tactic. A campaign, for instance, could state that the majority of voters support their party's nominee, despite the fact that this isn't precise and reliable.
- Emotional appeals: regularly uses emotive language to affect people ’s thoughts. Propaganda, for example, may utilize fear or fury to convince voters to support a specific idea.

Someone's propaganda is another one's truth, hence the subject of propaganda has generated controversy throughout history. As a result, there are several widely accepted generalizations concerning propaganda. To spread or support certain ideas is what is meant by propaganda, in the broadest and general sense. This shows that propaganda does not always have a negative perception, despite the fact that most of what we consider to be propaganda does.

Propaganda is divided into three categories: white, grey, and black (Jowett 2012). White propaganda is factual in most cases and comes from a reliable source. The material is displayed in a way that makes the source and sender look favourable. The information's veracity is uncertain when it comes to grey propaganda, and the source's identity might or might not be appropriately stated. Grey propaganda lies between black and white propaganda. The purpose of black propaganda is to propagate falsehoods, falsifications, and misrepresentations. Its source is either hidden or unreliable.

In the twenty-first century, a number of governments established pro-state propaganda networks on the Internet, which are not carried out by the media, but by individual users - commenters on sites with user content. Examples of these systems include "Kremlin bots" in Russia and "Umaodan" in China. The objectives of such clusters usually involve writing articles and messages for social network platforms, topics and discussions for web forums, and comments on online media sites, frequently taking the form of violent bullying of rivals.

The employment of simulation tools by a propaganda group to work online, which constantly generates millions of bogus messages from thousands or perhaps more phony virtual personas via the most prominent social networks, is one of the key indicators of their involvement. Nevertheless, this artificial crowd is what differentiates fraudulent accounts from actual people using the internet who have the freedom to express their personal opinions through the Internet. Propagandists use sophisticated software to intentionally eliminate real people from digital platforms, replacing them with fictitious public, and impose on community a vast deception concerning purported existence of a non-existent overwhelming public opinion on a certain subject of social life.

Since the perception is mainly left to the audience, propaganda has constantly been tough to define. While conventional methods of evaluating propaganda work in practice, the summary is mostly biased. Technology's development and the rise of social media have made this claim more relevant because anybody can spread whatever message they choose and potentially reach a huge audience. Twitter has been evolved as a venue for the transmission of ideas since messages can be created easily and quickly, disseminated to a large audience anonymously, and pose minimal risk to the author (Guarino et al. 2020).

3.1.3 Characteristics of propaganda on social media

Ellul (1965) used the terms vertical and horizontal propaganda as basis of a strict paradigm to differentiate both propaganda from social elites and from small citizen entities. Vertical propaganda, also known as propaganda of agitation, is a tactic used by elites to convince the public to actions. One-to-many conversations are essential in this context since they allow for the large-scale mobilisation of individuals to carry out the source's instructions.

On the other hand, integrational or horizontal propaganda aims to stabilize the community, to reunite and reinforce it (Ellul 1965). Both societal elites, such as authorities attempting to maintain social stability during periods of political upheaval, and general public may use kinds of propaganda. Narrow, independent organizations collaborating on the basis of a shared philosophy are the foundation of horizontal propaganda. It stands out since it is "not originating from the top" and comes within the community.

Nowadays, social media platforms are used for social connection – whether it is regarding to politics, culture, or daily life. As a result of this evolution, propagandists may now implement new forms of vertical and horizontal propaganda that users can evaluate and spread by commenting, liking, retweeting, and sharing. Social media significantly reduced the cost of producing digital content in terms of horizontal propaganda. Social media platforms are accessible to anybody who has a functional computer or smartphone with Wi-Fi connectivity, unlike websites, which needed the purchasing and maintaining of the domain. This has created a new opportunity for people and small organizations looking to manipulate others for their own needs. As a type of obfuscated propaganda, cloaked websites often promote material as severe and reliable while obscuring the website's source. While social network platforms like Facebook or Twitter focusing on user profiles and the presentation of personal networks, are especially well adapted for disingenuous types of propaganda. Properly designing a fake identity and upholding it via postings that are evaluated by user comments, likes, and shares helps to build the reliability and credibility of a disguised social media account. Thus, a user who "likes" or "friends" a fictitious account or page has the capability to assist in both its dissemination and approval. Taking into account these characteristics, it can be said that propaganda via social media relies on an identity (veiled) that is formed through a profile or page, a flow of unique material, and a persona that is constantly replicated and negotiated through relationships among posts and comments (Farkas et al. 2018).

Anonymous Danish propagandists encouraged users by creating fictitious Muslim profiles on Facebook in 2015, falsely claiming that Muslims were intending to murder and rape (non-Muslim) Danes (Farkas et al. 2017). Propagandists garnered over 20.000 comments from Facebook users in Denmark through 11 Facebook pages. With over 10.000 comments were left on the most commented-on page, that remained online for less than 4 days until Facebook banned it. Most individuals who responded to these instances of bogus disingenuous propaganda were violent toward the sites and Muslims in general. The pages became openly racist and cruel through offensive comments. The propagandists behind the fictitious identities were able to sustain total anonymity because to Facebook's setup (that gives page owners seemingly limitless privacy and security).

In the scope of vertical propaganda, major corporations consider taking benefit of the decentralized nature of social media by planning extensive campaigns that are, despite their challenge in being recognized as such, are nevertheless tough to spot. Social bots and the labour of so-called troll armies are two essential elements in this area. Entities would perhaps pay users to post content using their own social media accounts or via networks of fake profiles, which can be used as propaganda of agitation and unification as they look to strengthen their position by attacks on perceived rivals as well as the falsification of public support (Farkas et al. 2018).

3.2 Twitter

Since the dawn of the twenty-first century, many of social media networks have appeared, but only a few of them have succeeded in becoming well-known and remaining competitive. Twitter has consistently ranked among the top tier in terms of consumer and sponsor satisfaction levels and overall user statistics (Shepherd 2023).

Twitter introduced the globe to the ground-breaking ideology of microblogging by showing that messages don't have to be lengthy to be understood by people all over the world. You may accomplish it for only 280 characters.

The first tweet was sent in 2006, marking the beginning of short message transmission. This occurred on March 21 at 20:50 p.m. Pacific time. At this moment, Jack Dorsey sent the message: "just setting up my twttr." This was the start of the Twitter era. Overall, this social network has considerably more potential than is often realized. Initially, only postings were permitted. Later, the system was enhanced, and video and photo files started to be uploaded.

Applying the right assessment standards and metrics is necessary to estimate Twitter involvement. To define Twitter activity, several academics employed particular behavioural markers like the quantity of "likes" and "retweets". Audience excitement in a tweet may be expressed by clicking the "like" button. When people retweet a post, it shows that they have carefully considered whether to share it and are attempting to get the word out about the material they perceive to be newsworthy. Followers are more engaged and interact more when you use likes and retweets, which boosts connection.

According to research of Website Rating Team, Twitter is used by 83% of the world's leaders and they are part of this global community (WRS, 2022). As of February 16, 2022, pursuant to Daily Sabah’s article, Obama is the most widely known former or current president, with over 130 million followers. Current U.S. president Joe Biden and Former U.S. Secretary of State Hillary Clinton, both have more than 30 million followers on Twitter, sharing the next positions on the ranking after Narendra Modi, the Prime Minister of India, who has just almost 76 million subscribers. The honourable second place in this rating, surprisingly, is taken by the Former President of U.S. with roughly 88.7 million followers (until his Twitter account was banned in January 2021) – Donald Trump.

Any news of a global interest may now be spread instantly around the world. The news that stunned everyone recently is a real proof of this: The highest point in Africa now has high-speed internet access (Princewill 2022). “Today (16 November 2022) Up on Mount Kilimanjaro: I am hoisting high-speed INTERNET COMMUNICATIONS (BROADBAND) on the ROOF OF AFRICA. Tourists can now communicate worldwide from the summit of Mount Kilimanjaro. WE ARE GOING TO UHURU PEAK 5880 Meters Above Sea Level!#royaltourcompliment #royaltour” posted via his official Twitter account Tanzania's minister of information, communication and information technology, Nape Moses Nnauye.

Despite the fact that 1.3 billion people are on Twitter, only about 330 million seem to be – frequent users. However, 48 million of Twitter's profiles, according to forecasts, are – bots.

3.2.1 Twitter bots

On Twitter, bots are automatic or semi-automated programs that make use of the platform's standard features, such tweeting, retweeting, and uploading material. Twitter, in contrast to other social media platforms like LinkedIn, permits the usage of bots on its platform. For instance, in 2014 a Twitter bot (@everyword), created by poet and programmer Adam Parrish, fulfilled the challenge of tweeting every word in the English language in alphabetical order (The Guardian; 2014). It is possible to set up a bot for personal purposes with the use of third-party software, and Twitter's API enables automated postings. The bots are generally innocuous and openly identify themselves as bots.

However, there is a fraction of political Twitter accounts that are distinctive and are frequently alluded to as political propaganda bots. These are Twitter accounts that claim to be actual people, and their divisive tweets. Instead of genuinely producing their own material, they usually retweet others'. These accounts have occasionally been found to be retweeting false content or calling for rioting or other forms of disruption. Additionally, these accounts appear to be associated with the identical networks of many political propaganda bots, enabling them to disseminate information to real users on the network relatively fast. But they make a lot of effort to appear natural. In fact, several of these propaganda bots initially seem to be human-like accounts, implying that a portion of these bots may have come from compromised or bought and sold human accounts. Indiana University research published in 2017 showed that, on average, 9% to 15% of all regular Twitter users are social bots (Varol et al.; 2017).

It's challenging to estimate the exact number of bots that are active on Twitter. During the time that the United States was under Covid-19 stay-at-home restrictions, new research by Carnegie Mellon University revealed a spike in bot behaviour. According to the Carnegie Mellon analysis (2020), roughly half of the Twitter accounts appealing for America to reopen could have been generated by bots. Over 200 million tweets that mentioned the new COVID wave and date back to January 2020 were examined for this investigation. 41 out of 50, or 82%, retweeters were discovered to be automated.

3.2.2 The challenge of spotting bot accounts

Analyses of artificial accounts' personal contributions to the social platform ecosystem are surprisingly underrepresented in research discussing their identification. Texts sent and received on social networks indeed include a bunch of data, however only a small portion of that data is essential for describing how the system works, with most of the remaining data operating as void. To figure out which accounts, along with bots, play an important role to the efficient distribution of information, it is crucial to identify the important forms of engagement. In this regard, it is vital to evaluate the natural network's attributes to a suitable data set (Caldarelli et al. 2020).

The challenge in recognizing bots on a social media network like Twitter stems from the fact that there is no way to properly understand what a bot looks like. These narratives, unlike analytic data samples, have no ground truth or labelling. It is almost impossible to identify bots if there is no understanding of exactly how they look, and, paradoxically, it is not possible to understand exactly how bots look like, because they are difficult to be detect.

Certain criteria designating specific suspicious activity as escalating guarantee of bot profiles in order to bypass this paradox is going to be applied in this thesis. High-assurance bot accounts are typically characterized by (Johansen) actions like:

- tweeting several more times per hour during the day.
- automated short replies
- resemblance of the content across many accounts
- newly created accounts with the lack of profile description, pictures and so on
- promoting controversial political misinformation (particularly blatant propaganda).
- gaining a significant flow of following accounts extremely rapidly.
- quite often retweeting and boosting other suspicious profiles.

The above-mentioned parameters assist us in some instances in distinguishing these eye-catching bot accounts, but any straightforward principle method is going to contain some exclusions. For instance, a famous person who just managed to create a Twitter profile will quickly accrue a significant amount of followers. Another scenario could be during any Sport/eSport event, when a fervent fan might tweet about spectacular plays just every couple of minutes.

Twitter bots depend on secrecy. They may mimic genuine individuals by like your tweets and content while acting as artificial modules. Alternately, they could behave maliciously by using bots to harass, intimidate, convince, and inspire you to act in ways that are motivated by misleading info and to question things that aren't necessarily true.

Twitter bots have been used by cybercriminals to simultaneously disseminate malware-containing dangerous information to huge numbers of Twitter users. Avoid following links in tweets and other messages from unidentified or dubious sources to help defend yourself against such viruses.

3.2.3 Algorithmic amplification

In late 2021, Twitter's Staff Machine Learning Researcher Luca Belli shared with the public his report regarding the investigation of Twitter’s algorithmic boosting of political content. The first section of the research analysis tweets from elected politicians in seven countries (Canada, France, Germany, Japan, Spain, the United Kingdom, and the United States). Whereas tweets from political representatives represent only a small part of political material on the network, they investigated whether computer recommendation engines intensify information from other sources.

The outcome was unforeseen: some political information gets boosted on the platform. According to authors, these observable patterns are indeed a consequence of interactions among individuals and the system. Moreover, determining why they exist is a considerably major challenge to resolve.

The study was concentrated on the dynamic relationship of both an algorithmic engine and the network's users. Since every algorithm amplifies by default, algorithmic amplification is not an issue. If an algorithm receives special privileges based on the way it was built rather than the experiences users encounter with it, thereafter algorithmic amplification would be an issue (Belli 2021).

3.2.4 The importance of hashtags in Twitter

In the paper "#AdvocatingForChange: The Strategic Use of Hashtags in Social Media Advocacy", another research that employed hashtags to target Twitter users can be found. The National Health Council's 105 employees were tracked in this study for an 8-month interval for their usage of hashtags on Twitter (Saxton et al. 2015). This study sought to understand how advocacy groups interact with hashtags, what kinds of hashtags are being used, and if hashtags raise audience engagement. Eight distinct categories of hashtags were identified from the study's discovery of 9,934 distinctive hashtags used during that span.

Public education hashtags (50.4%), event hashtags (19.3%), call-to-action hashtags (3.2%), hashtags that reflect an organization's values and goals (9%), branding hashtags (7.2%), dialogic hashtags (5%), time and place hashtags (3.3%), and business hashtags (2.2%) are among the many types. The percentages show how frequently each category was used in data set that was gathered. The authors hypothesized that hashtags related to public education were the most popular because they may be used to enlighten constituents as well as to create a network of knowledgeable supporters who may support advocacy activities in the future. Another intriguing discovery was the positive correlation between the number of hashtags in a message and the number of retweets it gained, especially when several types of hashtags were used.

Three main hashtag usage tactics were noted by the authors:

- to enhance engagement – by using several hashtags to increase the amount of retweets.
- to identify deeper connection within the industry – by using hashtags that other individuals are using in their area of work.
- to use hashtags patterns that users feel would be more recognized by their audience.

The number of hashtags used is not as important as the sort of hashtag used during the process of generating interaction, it was discovered once there was an attempt to mix those tactics to determine which is more successful.

3.3 Russian activity

While Russian propaganda on Twitter has entirely failed with the Ukrainian audience, it still somehow efficiently reaches Western viewers. 7 million subscribers and more than 35 million retweets are just found on the accounts of the Russian government.

The social network has a significant impact on public life; it serves as a mechanism for both propaganda and charitable causes. For contrast, Sergei Pritula assisted the Ukrainian army earn 5.5 million UAH by tweeting about the Eurovision finals. At the same time, but activity increased throughout the contest, spreading Russian narratives regarding "Ukrainian Nazis" and requests to disqualify Kalush for making political remarks (Frost 2022).

Just because of the high proportion of accounts belonging to public authorities does Twitter come under fire for spreading political statements. An analysis on the performance of the platform's algorithms was posted in an article on the blog in 2021 (Belli 2021). It came out that the network frequently favours "right-wing" politicians' views over "left-wing" ones when promoting their ideas. Moreover, algorithms propose additional media content with a "right bias" in it. There is no official justification regarding this yet.

Despite having fewer subscribers than other social networks, Twitter has emerged as a crucial tool for creating and sharing fake news about the Russian invasion to Ukraine.

Journalists from The Conversation investigated the activities of 75 –official Russian government accounts. They have received over 35 million retweets – and over 7 million subscribers combined. From February 25 to March 3, 2022, –these accounts sent 1,157 tweets, over three-quarters of which were hoaxes –concerning the invasion of Ukraine and attempts to clarify it (Graham et al. 2022).

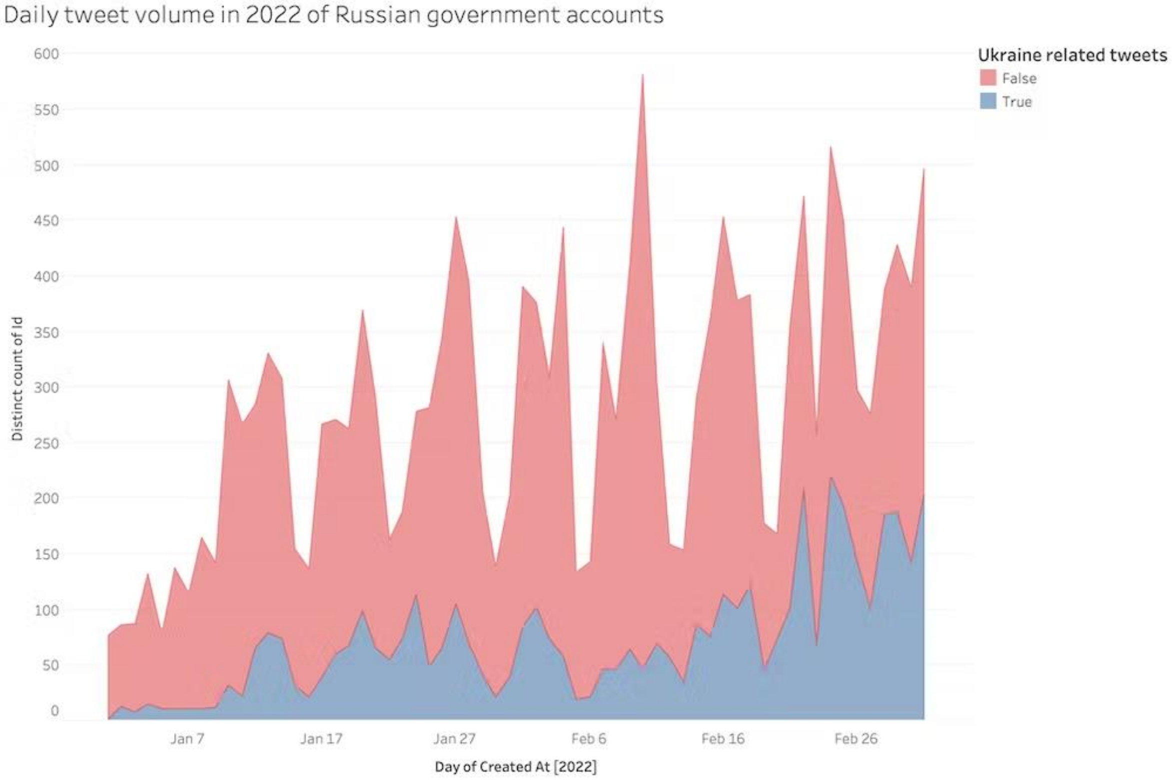


Figure 1 - Daily tweet volume in 2022 of Russian government accounts (coloured by tweets about Ukraine versus other).

The Economist (2022) also investigated how propaganda affected the conflict between Russia and Ukraine. They examined 3.7 million tweets from 7756 accounts for pro-Russian propaganda and geographic orientation using data from studies from the British analysis company CASM Technology. The hashtags # IStandWithPutin and #IStandWithRussia were used to identify them. It turns out that propaganda is quite powerful in Turkey and India, and during the first 12 days of the invasion, there was a disproportionate amount of activity using doubtful accounts. 56 percent of them are still actively disseminating misinformation.

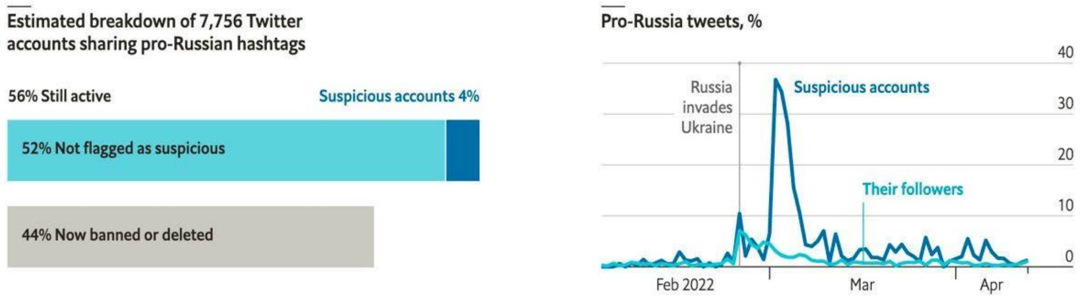


Figure 2 - Source: The Economist.

Evenmore, the semantics and phrasing of the tweets were also examined by The Economist. They discovered a lot of accounts that, in contrast to bots, compose unique messages using diverse phrases. These are most likely actual individuals who receive compensation from Russia or its allies.

3.4 Chinese Activity

The Chinese government's narratives about what was occurring in Xinjiang were being promoted by around 2,000 Twitter pages, a majority of which displayed anti-Western attitude or dismissed the claims made against China, according to findings conducted by the Australian Strategic Policy Institute in 2021 (Ryan et al. 2021).

China frequently integrates social media to disseminate its perspectives, and research conducted in 2021 by Associated Press and the Oxford Internet Institute (Schliebs et al. 2021) revealed that vast numbers of fake profiles amplified disinformation by Chinese officials and state media enormously many times to attract a broad audience despite covering the obvious truth that the material is governmentally.

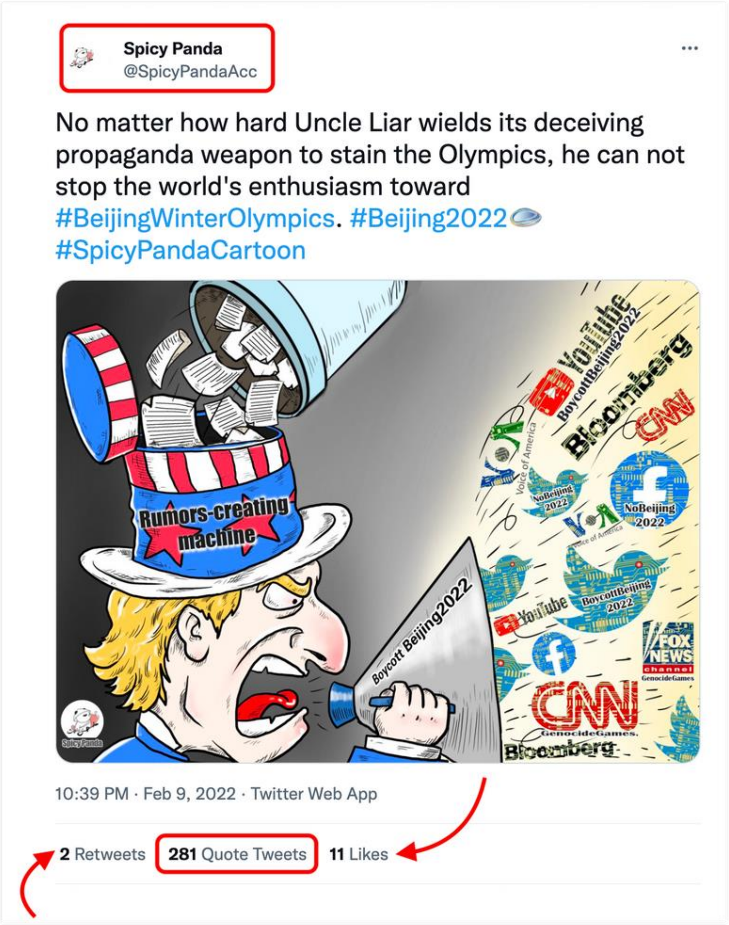
In order to promote the Beijing Winter Olympics on the social media networks such as Instagram and TikTok, China initiated a secret government marketing campaign in 2022. The country hired a U.S.-based company to find bloggers in the western nation (Associated Press 2022).

3.4.1 Olympics Games 2022

Despite having solid dominance over what audiences within its own nation absorb, China has used a wide range of digital platforms to disperse the exclusive interpretation of the Beijing 2022 Winter Olympics outside of its boundaries. This has given China's perspective, in many ways, higher engagement and subtext than before.

China has managed to manipulate the way the actions have been portrayed, even outside of the nation, using bots, fictitious profiles, real individuals, and other methods. They have promoted all that supports the government, upbeat narrative about the Beijing Olympics and attempted to limit things that disagrees with (Mozur et al. 2022).

Furthermore, China has made covert attempts to sway internet debates. By exchanging local news posts with exactly similar messages, The New York Times and ProPublica discovered a network of almost around 3,000 phony Twitter profiles that seemed to be collaborating to boost the Olympics. These accounts claimed getting newly generated with a small number of subscribers, regularly retweeting rather than offering original content, and looked to exist only to boost official Chinese viewpoints.



Spicy Panda
@SpicyPandaAcc

No matter how hard Uncle Liar wields its deceiving propaganda weapon to stain the Olympics, he can not stop the world's enthusiasm toward #BeijingWinterOlympics. #Beijing2022
#SpicyPandaCartoon

10:39 PM · Feb 9, 2022 · Twitter Web App

2 Retweets 281 Quote Tweets 11 Likes

Figure 3 - SpicyPandaCartoon

They have focused enough of their attention on the Spicy Panda profile, that has been publishing caricatures and recordings to fight requests for an Olympics boycott. Spicy Panda's followers were found to be represented by 861 accounts, 90% of those were registered after December 1.

Dozens of those same profiles revealed by the Times and ProPublica had their accounts terminated, Twitter announced via an email, for breaking the company's anti-spam and anti-platform abuse rules. They stated that it continued to look further into connections between the accounts and options for influencingsupported by the government.

3.4.2 Human rights abuse

Regarding how the Uyghur ethnic minority in Xinjiang is treated, the Chinese Communist Party (CCP) has drawn criticism from throughout the world. The Xinjiang Uyghur Autonomous Region and the exploitation of Uyghur labour exploitation in Xinjiang are the crucial points that Nisos report (2022) conducted to be the core of an organized, bogus network of 648 Twitter accounts, certain of which dates to the early November 2021. The collaborative misleading network tweets, according to Nisos researchers, are seeking to simultaneously obscure critical Twitter commentary about Xinjiang and advance a favourable thesis about the status of Xinjiang and Uyghurs in the People's Republic of China (PRC). An international crowd is the action's primary objective.

The fake news story of freedom and a prosperous way of life for the Uyghur community in Xinjiang is being promoted through a planned, dishonest network. Additionally, it is employed to refute claims made by outsiders that the local government has subjected them to domestic workers. Pleasant tweets and video content promoting life in Xinjiang along with others that contradict the existence of slave labour there make up the falsified network activity.

About 2,803 unique tweets were found by Nisos investigators among the 648 fake accounts they had previously detected. A total of 773 quotations and 1,136 retweets by the 648 profiles were also found by Nisos analysts (2022).

The bulk of tweets were published around 9 a.m. to 7 p.m. in China, within office hours, according to Nisos analysts, who didn't name the people who are responsible for the subnet of fake accounts. Most of those profiles have been registered after August 2021, and those profile photos were default pictures. The messages were mostly sent minutes apart.

To increase their exposure on the site, the accounts constantly referenced other profiles on the network, while periodically they would also emphasize posts of Chinese officials like Zhao Lijian, a representative for the Chinese Foreign Ministry, and Zhang Meifang, the consul general of China in Belfast, stated Nisos (2022). In response to violations of Twitter policies, several of those Twitter accounts listed in that research have already been banned.

3.5 COVID-19 Pandemic

Based on Wikipedia, the free Encyclopedia, the COVID-19 pandemic, sometimes referred to as the coronavirus pandemic, is a persistent worldwide outbreak of the coronavirus illness 2019 (COVID-19), which is brought on by severe acute respiratory syndrome coronavirus 2. (SARS-CoV-2). In December 2019, a viral epidemic in Wuhan, China, resulted in the detection of the new pathogen. There were attempts to contain it, but they were unsuccessful, causing the infection can spread to many other regions of China and eventually to the rest of the world. The epidemic was deemed a worldwide public health crisis on January 30, 2020, and a pandemic on March 11, 2020, by the World Health Organization [WHO]. The pandemic, which was one of the worst in history as of August 2022, has resulted in more than 597 million people were infected and 6.45 million fatalities that were reported.

To manage the outbreak, mankind still requires an effective vaccine, despite the fact that almost all governments throughout the world have developed and implemented a wide range of public health and social restrictions, including mobility restrictions, social isolation, and individual measures. Recently, the emphasis of interest on a global scale has been on the investigation and creation of COVID-19 vaccinations. Russia had its COVID-19 vaccine certified for sale as early as August 11, 2020, but while widespread vaccination certification didn't even start till December 2020. By the end of 2020, the COVID-19 vaccination has been licensed for use in the UK (2 December), the US (11 December), the EU (21 December), China (31 December), and other nations and regions (Wang et al. 2022).

According to prior studies, obtaining the COVID-19 vaccine is contentious issues across many countries. Whereas others are concerned about potential negative effects, some individuals question the vaccination's efficiency. The public's unwillingness to get the COVID-19 vaccine and mistrust in the vaccination campaign, according to Cornwall, has grown as a result of a number of reports about adverse effects published on social media. Merely 71.5 percent of the surveyed individuals from 19 different nations or geographic areas mentioned that they intended to receive the COVID-19 vaccine, according to study by Lazarus and co-workers, despite the vaccine's shown effectiveness and safety (Wang et al. 2022).

During the battle for survival against COVID-19, it's crucial to reduce vaccine scepticism and encourage vaccinations. The news media could be a major factor in the conceptual production of existence. Reports regarding the COVID-19 vaccine from leading news agencies, the world's largest broadcasters of information, have a significant impact on how the public views immunizations. According to research, foreign media organizations have the authority to regulate both the journalistic objective of paper and digital news outlets. Among the several foreign news organizations, the main three, the Associated Press (AP), Reuters (Reuter Ltd.), and Agence France-Presse (AFP), have always given their readers unbiased information (Wang et al. 2022).

These companies have successfully started using Twitter to reach their followers. Investigators noted that authors frequently do not pay that much to the wants and tastes of the audience and that Twitter is regularly utilized as a one-way communication system. According to analysts, senders should enhance their social media content to increase customer loyalty if they wish to fully benefit from Twitter's potential benefits. Understanding concerning healthcare, a sense of identity and socialization might all be managed to improve via audience interaction.

A brand-new hashtag, #FilmYourHospital, makes its debut on Twitter in March of 2020. In order to demonstrate that the COVID-19 epidemic is a sophisticated fake, the hashtag urged individuals to check out nearby ambulances and shoot photos and videos of unoccupied hospitals. These influential users promoted the initiative using this hashtag, inspiring their readers to breach the embargo and record what was occurring in their neighbourhood facilities. The spread of the #FilmYourHospital hoax from a single tweet highlights the persistent difficulty in countering misleading, spreading information since the COVID-19 outbreak (Gruzd et al. 2020).

Reasoning and pointing individuals to reliable global health agency media sources might possibly slow the spread of disinformation, but it is significantly more challenging to disprove unfounded and inaccurate statements that are motivated by politics and backed by strong principles rather than scientific evidence.

3.6 Methods of identifying fake accounts on Twitter

Researchers created a variety of techniques to identify malware Social Media Bots (SMBs) as individual accounts or campaigns to prevent the spreading of these destructive SMBs. These techniques were initially divided into graph-based, feature-based, crowdsourcing-based, and mixed approaches by Ferrara et al. (2016). With the addition of subclasses beneath the primary categories, this classification system was improved by Adewole et al. (2017).

The taxonomy in this research is created in a way where it will include all currently used approaches. As the former taxonomies were unable to define various approaches used in studies, new categories and further subcategories are implemented by Orabi et al.(2020). The resulting classification is shown in Fig. 3.

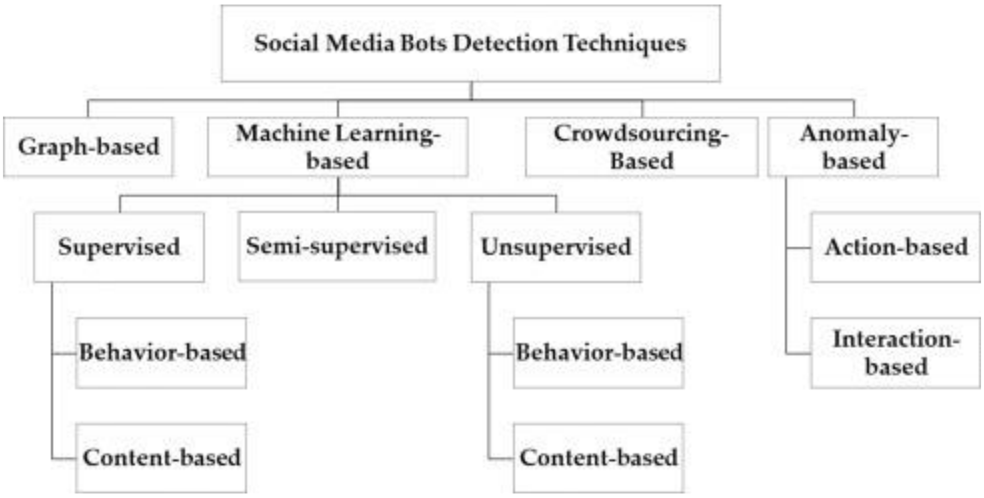


Figure 4 - Proposed taxonomy for social media bots detection methods. (Orabi et al. 2020)

Graph based. A mathematical graph is a set of points and line segments that are commonly used to represent pairwise relations between objects, such as in social network structures. Graph-based methods have been used to detect bots in online social media by leveraging the properties of social graphs. Several approaches have been proposed, including Integro (Boshmaf et al., 2016), which uses supervised machine learning to rank real users higher than fake users based on the network structure of OSN. Another approach by Cornelissen et al. (2018) combined network structure measures and unsupervised machine learning methods. Hurtado et al. (2019) exploited temporal and network information to detect political bots on Reddit, while Yang et al. (2014) used an interaction graph model in a supervised classification approach. Dorri et al. (2018) relied on the homophily property of social network graphs to detect bots using semi-supervised machine learning. BotCamp (Abu-El-Rub & Mueen, 2019) and Ahmad and Abulaish (2013) also used graph-based methods to model topographical information and social media profiles, respectively, to detect bots through clustering.

Machine learning based. Machine learning is among the most reliable and fairly priced methods for handling huge data-related issues. Buczak and Guven (2015) cite Arthur Samuel's definition of machine learning as the "area of research that offers machines the capacity to learn rather than being explicitly programmed" (2015). Machine learning techniques draw information from analytical findings and experience. Due to these benefits, machine learning has a wide range of applications in a variety of fields, such as business (Bose & Mahapatra, 2001), education (Kahraman et al. 2010), biological and medical research (Libbrecht & Noble, 2015), and cybersecurity, which is the subject of our investigation.

The large number of the material that methods turned up used machine learning. Indeed, a study on methods for detecting bogus accounts also revealed same result (Adewole et al., 2017). This data demonstrates the critical role machine learning plays in identifying illegitimate accounts and bots on social media.

Supervised approach. In all types of detection methods, supervised learning-based detection techniques are the most prevalent. The objective is to create a model of the predictive feature-based distribution of class labels. The resultant classifier then forecasts class labels for cases that are not yet classified given values for the prediction characteristics (Kotsiantis et al. 2007). In other terms, by applying training data on a collection of labelled accounts, a classifier learns how to recognize accounts (as bots or not) depending on a typical pattern of accounts traits. To develop the model and create the classifier, a dataset of labelled cases and their obtained features is required. The importance of the set of differentiating characteristics and the effectiveness of the training set affects how well a model performs.

Two main categories of bot traits are exploited. The first category is behavioural characteristics, which represent users' data (metadata), activities, interactions, timestamps, and maybe even text counts without in-depth textual content analysis. The second is content features, which examine users' textual content to separate bots from actual people. The majority of supervised methods used in articles may be classed as behaviour-based strategies, which focus only on exploiting behavioural aspects. Moreover, there are strategies known as content-based techniques that primarily employ content elements, occasionally in combination with behavioural aspects.

Behaviour-based. Various researchers have developed methods to identify bots on Twitter. One such method is BotOrNot (Davis et al. 2016), which measures the probability of an account being a bot. Other methods include a Chrome browser plug-in by Alarifi et al. (2016), which detects if an account is human, Sybil or Cyborg. Kantepe and Ganiz (2017) used the most efficient features from DARPA competition to classify Twitter accounts as normal or bot accounts. David et al. (2016) identified features to detect Sybils and used five different classification approaches, with Random Forest Classifier performing the best. Velayutham and Tiwari (2017) proposed BotClassifier, which shows better classification results than Naiive Bayes classifier. Another method, CATS (Amleshwaram et al. 2013), uses entropy, community nature of spammers and a blacklist of URLs used by spammers to detect Spam bots on Twitter.

Ji et al. (2016) introduced features to detect social bots and evasion mechanisms. Teljstedt et al. (2015) proposed a semi-automatic approach to detect bots on Twitter. Gilani et al. (2017) partitioned Twitter accounts based on popularity and identified efficient features for classification. Daouadi et al. (2019) proposed an enhanced set of features based on the amount of interaction of an account and how much other users interact with it to detect automated accounts on Twitter. Yang et al. (2014) detect bots in marketing campaigns using users’ interactions. Fazil and Abulaish (2018) proposed community-based features to detect Twitter Spam bots. Chu et al. (2012) classified Twitter accounts into humans, bots, and Cyborgs using an entropy component, a machine-learning component, an account properties component, and a decision-maker component.

Content based. Numerous articles discuss the use of content analysis and textual information for detecting Twitter bots. Kudugunta and Ferrara (2018) used deep learning and six account features to identify bots, and oversampling techniques were used to improve the dataset. Wang et al. (2018b) used tweet similarity to detect social bots, while Ping and Qin (2018) used tweet content and metadata with a CNN-LSTM algorithm to extract user information. Morstatter et al. (2016) used BoostOR to classify accounts based on their post topics, and Igawa et al. (2016) used a wavelet-based approach with a Multilayer Perceptron and Random Forest Classifier to classify accounts into human, bot, and cyborg.

Bara et al. (2015) built a model to detect Twitter Spam bots based on tweeting pattern similarity and proximity to seeded Spam. Budania and Singh (2017) classified Twitter users into person and non-person, and Dickerson et al. (2014) introduced sentiment-based features to detect bots. Loyola-González et al. (2019) used sentiment analysis to build a model that uses Contrast Pattern-Based classifier to detect Twitter bots. Finally, Beskow and Carley (2019) classified usernames into random and non-random and collected a dataset of 235,000 Twitter accounts with random usernames to enhance the detection of bots on social media.

Unsupervised approach. Unsupervised machine learning is a method where the algorithm clusters the input data (El Naqa & Murphy, 2015). To put it another way, the unsupervised approaches concentrate more on what is prevalent to groups of accounts and nodes accounts according to similarities among accounts in a single group, eliminating the need for labelled data to identify bots using these methods. They also rely less on the values of various characteristics to define each account. The suggested taxonomy divides articles that apply unsupervised learning techniques into behaviour-based and content-based categories, almost as articles that have used supervised learning methods.

Behaviour-based. Two unsupervised machine learning models, DeBot and Digital DNA, are employed to find social media bots. Chavoshi, Hamooni, and Mueen (2016a, 2016b) created DeBot, which leverages connected activity among groups of accounts to identify bots on Twitter. If an account tweets at least forty tweets within an hour and has highly linked activity across groups of accounts, the model labels the account as a bot. The chain of online activity for a social media account is encoded by the Digital DNA model, which was put out by Cresci et al. (2016). The program analyses the digital DNA fingerprints of Twitter accounts and supposes that campaigns run by spam bots are groupings that share similar action sequences (Longest Common String).

Content based. To spread their message, bots must be visible and replicate content. Chew (2018) used these assumptions to detect patterns of similarities and detect automated Twitter accounts, which appear to be influence bots. Chen et al. (2017) also exploited content similarity and URL shortening services to detect spam bot campaigns on Twitter. They developed a system that monitors the top trending URLs in tweets on Twitter's real-time streaming and flags accounts as bots if they share similar recent tweets. The system then maps each campaign to the registrant email of the URL that the campaign shared.

Abu-El-Rub & Mueen (2019) used content to detect social media bots in BotCamp by exploiting trending topics to identify political discussions. They used DeBot to detect synchronized bots that hijack trending hashtags and then used graph-based methods to cluster the collected bots based on the graphs. Finally, a supervised model classified user interactions to identify bot campaigns in political discussions.

Semi-supervised approach. Since it employs partly labelled data, semi-supervised machine learning lies between supervised and unsupervised machine learning. This means that techniques of this type employ a significant amount of unlabelled data and a fraction of labelled data when creating classifiers in order to lower the expense of gathering labelled examples and elevate the classification precision (Zhu, 2005).

Shi et al. (2019) suggested clickstream sequences and applied semi-supervised cluster analysis to identify malicious social bots as a comprehensive characteristic that can't be commonly reproduced by bots. The authors claim that clickstream sequences can both reveal key aspects of a user's behaviour while also leading to a shift in that behaviour. In order to identify spam bots on the Twitter platform, Dorri et al. (2018) introduced SocialBotHunter, a model that makes use of user social behaviour and interactions. This model is based on the interpersonal attraction feature of social network graphs. By including a dataset that contains a sample with only labelled authentic users, the model can operate.

Crowdsourcing-based. A group of individuals are requested to complete a manual work as part of a kind of internet-based activity known as crowdsourcing (Estellés-Arolas & González-Ladrón-De-Guevara, 2012). This approach has a significant time or financial cost when used to find social media bots. So far, several investigators employed this approach to gather labelled datasets in order to conduct following study.

Wang et al. completed one of the key parts of research in this field (Wang et al., 2013). Wang et al. used a social Turing test in order to verify if a user of an online social network could accurately identify a user as Sybil. In order to improve the precision of Sybil identification, the authors developed a method that screens profiles using computer algorithms and therefore produces suspect accounts, and those would be further researched by chosen crowd workers. Also, 10 specially selected and qualified volunteers who performed as crowd workers to personally categorize Twitter profiles as humans, Sybils, and Cyborgs were used by Alarifi et al. (2016) to create their ground truth database.

Moreover, Cresci et al. (2017a) examined the effectiveness of crowdsourcing for identifying social media bots. The findings indicate that while crowd workers were successful in identifying legitimate accounts and classical spam bots, they were unable in identifying social spam bots.

Anomaly based. This class of detection methods relies on the presumption that genuine members of Online Social Networks wouldn't be motivated to exhibit any extraordinary behaviours because doing so would not benefit them in any way. As a result, it is quite probable that a user who acts suspiciously in specific circumstances is malevolent. A group is labelled a social botnet once abnormal group behaviour is observed. Often, the first stage of bot identification uses this technique. In order to improve future classification technique, ground truth datasets will be built, and information will be gathered from the discovered bots.

Action-based analysis refers to several systems that identify unusual account behaviour without engaging the corresponding accounts. Some methods, known as interaction-based identification, rely on setting off the abnormal behaviour of bot accounts and spotting them through their unusual interactions with a researcher-set trap.

Action based. Echeverria and Zhou (2017) randomly selected a sample of 6 million English-speaking Twitter users and found an unexpected pattern in the locations of a subset of individuals' tweets. This group had several distinctive features, notably posting spontaneous sentences from the "Star Wars" book. The researchers decided to label this group as bots as a result of their unusual actions. The following assumptions were made by Pan et al. (2016) according to the time-based patterns of posting attitudes:

- There is wide difference in human behaviour.
- The action of a bot is relatively easy compared to a real person.

A bot will therefore have low volatility. The investigators classify the accounts on Sina Weibo among human, bot, and cyborg categories using abnormal burstiness factor and time-interval evaluation score. One of their noteworthy results was that human accounts post significantly more frequently than bots do.

Interaction based. Attempting to broaden their communications infrastructure, Lee et al. (2011) relied on the unusual behaviour of fraudulent accounts that involved engaging with 'uninteresting' profiles. Investigators found viral abusers on the Twitter platform by using the honeynet method. 60 honeypots were made by the researchers to serve as conventional content violators. They labelled the accounts that engaged with the honeypots as content offenders, reasoning that a real user wouldn't be attracted to contact or follow them. The similar approach was used by Morstatter et al. (2016) to create the Arabic Honeypot Dataset, a freely accessible set of data.

3.7 Comparison of positive and negative-leaning tweet share

One of latest study on how a new corporation's political position affects the attitude displayed and the amount of interaction it receives on social media evaluated at 140.358 tweets from 24 left-leaning and 24 right-leaning news outlets, split according to Allsides' media bias grading system (Bellovary et al. 2021). The purpose of the research was to explore at the emotive content that mainstream media outlets tweet and the interactions that occur between users and that content. The information was gathered during a period when issues like the Covid-19 outbreak might have overwhelmed public attention on both political sides. Considering this scenario, the findings showed that news sources with political leanings on the left and right showed significantly negative instead of positive reaction on Twitter. The research suggested that while social media is generally beneficial, media companies' data could be an exception. This analysis reveals that news organizations with both left- and right-leaning ideologies may benefit from the cliche "if it bleeds, it leads" on social media.

It is significant that no variations in the representation of positive or negative affect across news sources with a left- or right-leaning leanings were discovered. Considering this, the study's findings lend credence to the idea that groups with a left or right leaning may not always exhibit affective expressions that are fundamentally different from one another. Additionally, the study contends that for media sources with both a left and a right-leaning, negative news material engages users more strongly than positive news material. As a result, for both right- and left-leaning political networks, negative news spreads more quickly within the current Twitter network than positive news. The results might be linked to the fact that material with a negative connotation has a stronger effect on behaviour and interest versus information with a positive meaning.

3.8 Summary of the key findings

The examination of earlier works on the subject of propaganda in general, and its propagation on the Internet in particular, helped the author of this thesis to realize that this is a highly relevant and significant issue on a worldwide level. Information has been shared for progression from the beginning of civilization with the goal of bettering everyone's quality of life. However, since the word "falsification" has been coined, progress has assisted some people attain their objectives. This, together with the divergent political ideologies of other nations, made things worse and gave rise to the term "propaganda." Propaganda, in general, is a crucial instrument in controlling the public's perception, and it also serves as an offensive and defensive weapon in information warfare. Publicly available media like radio and television are regarded as the initial providers of false information.

But with the invention of the Internet, the geopolitical game gained tremendous momentum. The 2016 US presidential election may serve as evidence of this. Then, suddenly, a number of nations that appeared to have the least interest in this (such as Russia, China, etc.) turned out to be responsible for the election's pivotal moment, which led to the election of D. Trump. It is important to note that the in the beginning of 2021 Twitter organization indicated that they will eventually block the account of US President Donald Trump due to concerns that it would be used to promote violence.

Negative news does, however, tend to travel more quickly, according to a recent study on the issue of information distribution on Twitter, either positive and negative. This is due to a large variation in how people psychologically interpret information. At the outset of the Russian invasion, confirmation may be a daily update of information.

Naturally, a lesser analogue may be the first waves of the coronavirus epidemic, which have just lately overwhelmed humanity. Which, interestingly, diminished just as quick as they emerged. The difference is that in this instance, there were (at least) two camps: those who supported common sense and sought to obey the recommendations of the health care industry, and those who believed in existence of conspiracy theories and refused to go along with the government. While some attempted to cure the illness, others made an effort to disprove its presence.

Orabi et al. (2020) discuss the various methods used to detect social media bots. The authors provide an overview of four main categories of detection methods: behavioural, network-based, content-based, and hybrid approaches. Behavioural methods analyse patterns of behaviour, such as posting frequency and time of day, to distinguish between human and bot accounts. Network-based methods look at the connections between accounts and the network structure to identify bots. Content-based methods analyse the language, tone, and sentiment of social media posts to detect bots. Hybrid approaches combine two or more of these methods for improved accuracy. The authors conclude that no single detection method is foolproof, and a combination of approaches is necessary for effective bot detection in social media

3.9 Research question formulation

This research is based on the fact that in the era of technology development, digital political propaganda is used as a tool for influencing various layers of society, and the existence of entire organizations that actively use bots to spread the same misinformation to every user throughout the Internet, and at the same time keep going unpunished.

From this, the following questions are formed:

- ✓ What are the characteristics of fake accounts on Twitter?

In order to determine whether and how propaganda has evolved since the invention of Twitter – the author will undertake research based on this question.

4 Practical Part

Starting with the fundamental research strategy and objective, this part gives a general summary of the methodology's basis. Furthermore, in accordance with the investigation's aim and concept, the data collecting technique, rate, and choice were assessed. As well, the type of data used, and its sources are clarified in this section. Also, a complete explanation of the procedures and instruments used to evaluate the data is provided. This chapter offers the foundation for the authenticity and reliability of the study thereby briefly explaining how the key questions were answered.

4.1 Research Philosophy and Design

4.1.1 Research Philosophy

As a research philosophy of this work, author adheres to positivism in nature. Positivism is a philosophical approach that has played a significant role in shaping our understanding of the social world (Popper, 1959). At its core, positivism emphasizes the use of scientific methods to study social phenomena, with the goal of discovering objective truths about the world (Babbie, 2013). Positivists believe that social phenomena can be studied in the same way as natural phenomena, using quantitative research methods and statistical analysis to measure and analyse data (Walliman, 2017). This approach has been influential in many fields, including sociology, psychology, and economics, and has contributed to the development of many important scientific theories and models.

One of the key strengths of positivism is its emphasis on objectivity and the use of empirical evidence to support claims about the world (Babbie, 2013). Positivists believe that knowledge can only be acquired through empirical observation, and that scientific methods are the best way to discover objective truths about the world (Popper, 1959). This approach has helped to establish the social sciences as a legitimate field of inquiry, with methods and standards of rigor that are comparable to those of the natural sciences (Babbie, 2013). The emphasis on quantitative methods and statistical analysis has also helped to develop a robust and reliable body of knowledge about social phenomena (Walliman, 2017).

Despite its many strengths, positivism has also been criticized for its limitations (Guba & Lincoln, 2005). Some critics argue that the positivist approach neglects the subjective experience of individuals and the role of social context in shaping human behaviour (Babbie, 2013). Others argue that the emphasis on quantitative methods and statistical analysis can oversimplify complex social phenomena, and that some aspects of social reality cannot be easily measured or quantified (Creswell, 2014). Nonetheless, positivism remains an important philosophical approach for understanding social phenomena and has played a significant role in shaping our understanding of the world (Babbie, 2013). Its strengths and limitations continue to be debated by scholars and researchers in the social sciences (Guba & Lincoln, 2005).

4.1.2 Conceptual framework

The investigation relies on a binomial relationship between the bot score number (as a result of Botometer® algorithm by OSoMe) of accounts participating in posting, referred to responsive variable(s) and the keywords or hashtags used in that tweets of certain events (election campaigns, COVID-19, war), which are the explanatory variable(s). It is worth nothing that algorithmic analysis of Twitter profiles Botometer® is revealed on the basis of several key assessment criteria, including so called "spammer" and "fake follower" rates.

Considering the main research question regarding the characteristics of fake accounts on Twitter, the following hypotheses flow from this investigation:

Main Hypothesis:

$H_{0:}$ There is no relationship between the spammer/fake follower score(s) and the account bot scores.

$H_{1:}$ There is a relationship between the spammer/fake follower score(s) and the account bot scores.

Alternative Hypothesis:

$H_{0:}$ There is no correlation between the spammer/fake follower score(s) and the account bot scores.

$H_{1:}$ There is a corelation between the spammer/fake follower score(s) and the account bot scores.

4.1.3 Research Design

Since the research investigated the relationship between Twitter hashtags of certain events and the accounts engaged in tweeting/retweeting content over a ten (10) month period, time-series data sets were used for the analysis. More specifically, the timeframes for the US presidential elections from May 2016 to March 2017, the COVID-19 pandemic from March 2020 to January 2021, and the military activities of Russia in Ukraine from February 2022 to December 2022. As it was previously noted, this study conducted an explanatory study inquiry because time-series data were employed in the research as well as being intended for that purpose. In this aspect, time-series data enables the examination of a significant enough sample over a predetermined period of time. The availability of data on Twitter hashtags and keywords as well as user activity through the Twitter source also made time-series data analysis possible despite the time limitations.

4.2 Sample size, Data Choice, and Collection

4.2.1 Sample size

Sample size is a crucial aspect of research design as it directly impacts the validity and reliability of study findings. According to Polit and Beck (2021), "sample size refers to the number of participants in a study" (p. 364). A sample size that is too small may lead to underpowered analyses, resulting in inaccurate or inconclusive findings. On the other hand, a sample size that is too large may lead to a waste of resources and time without improving the quality of the results. Therefore, selecting an appropriate sample size is crucial in research design.

Several factors influence the determination of an appropriate sample size, including the research question, the level of precision required, and the variability of the outcome measures. According to Tabachnick and Fidell (2019), "there is no simple formula to determine sample size, as the appropriate sample size depends on the specific research question and design" (p. 117). For instance, a study with a narrow research question may require a smaller sample size than a study with a broad research question. Additionally, a study that aims to detect small effect sizes may require a larger sample size than a study that aims to detect large effect sizes.

The author used primary data including the hashtags/keywords of specific events on Twitter. For each individual case, it was freely accessible and pulled from the Twitter social network over a 10-month period, from May 2016 to December 2022. It's indeed sufficient to claim that throughout these periods, a sample size large enough to perform statistical tests including linear regression, correlation analysis, and other statistical methods.

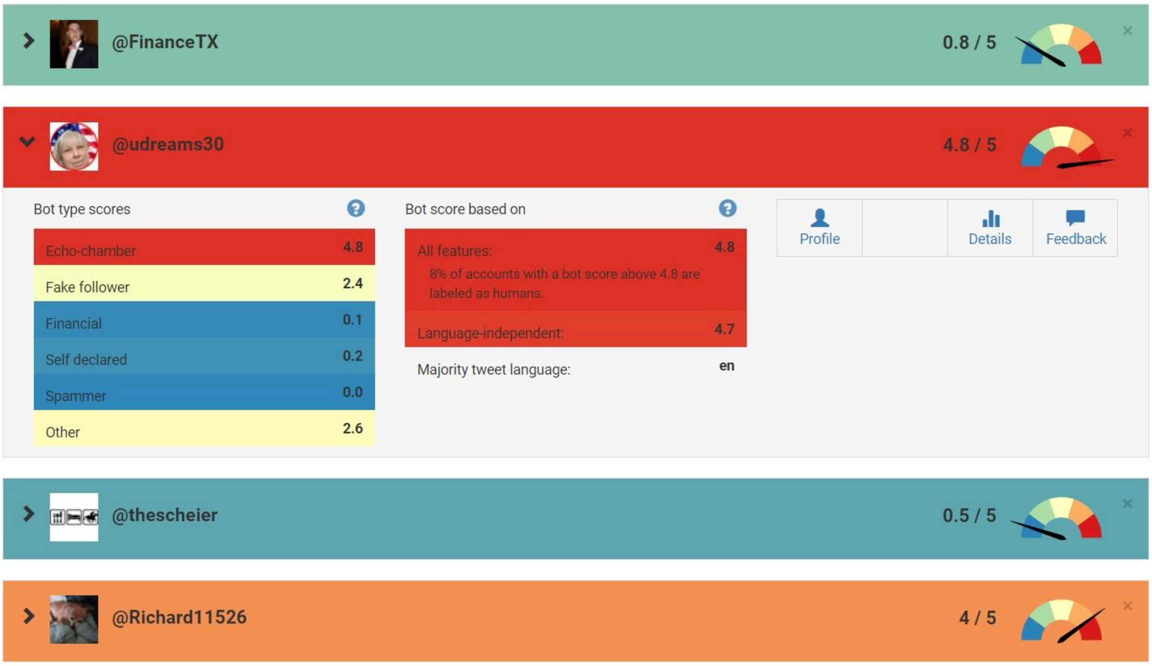
4.2.2 Type of Data

Based on above mentioned bot score statistic, the foundational dataset of this work is going to be a quantitative data type. Quantitative data type is a type of research data that is collected through numerical measurements and statistical analysis. This data type provides a structured and objective approach to research, enabling researchers to test hypotheses and make predictions based on statistical analysis of the collected data (Creswell, 2014). Quantitative data can be collected through surveys, experiments, and other forms of structured data collection methods, and can be analysed using various statistical techniques such as correlation analysis, regression analysis, and hypothesis testing. One of the strengths of quantitative data is its ability to produce objective and reliable results that can be replicated in other studies, providing a strong foundation for scientific research (Babbie, 2013). However, it is important to acknowledge the limitations of this data type, including the potential for oversimplification of complex phenomena and the potential to miss important qualitative data that may be difficult to quantify (Guba & Lincoln, 2005).

Author decided to choose a primary data as a source for further investigation mainly because it is unique and original dataset that has not been analysed or interpreted by anyone else. This means that the data collected is specifically tailored to research questions, providing a high level of relevance and accuracy to the data collected. Collecting primary data will enable to answer research questions directly, as author will have full control over the data collection process and can ensure that the data collected is of high quality. Additionally, it makes possible to ensure that the data is up-to-date and relevant to current trends, enabling to draw more accurate conclusions from analysis. While collecting primary data can be time-consuming and may require additional resources, the benefits of collecting original data outweigh the challenges, as it provides a more comprehensive and accurate understanding of the research question being investigated.

4.2.3 Data Collection

Initially, even at the stage of preparing the theoretical part of this work, the author came across a research work (Davis et al. 2016), which used the BotOrNot algorithmic system (current Botometer®) to analyse more than a million requests for data obtained from Twitter. The Observatory on Social Media (OSoMe) and the Network Science Institute (IUNI) of Indiana University collaborated on the Botometer project, which rates a Twitter accounts based on its activity.



@FinanceTX 0.8 / 5

@udreams30 4.8 / 5

Bot type scores	
Echo-chamber	4.8
Fake follower	2.4
Financial	0.1
Self declared	0.2
Spammer	0.0
Other	2.6

Bot score based on

All features: 4.8
8% of accounts with a bot score above 4.8 are labeled as humans.

Language-independent: 4.7

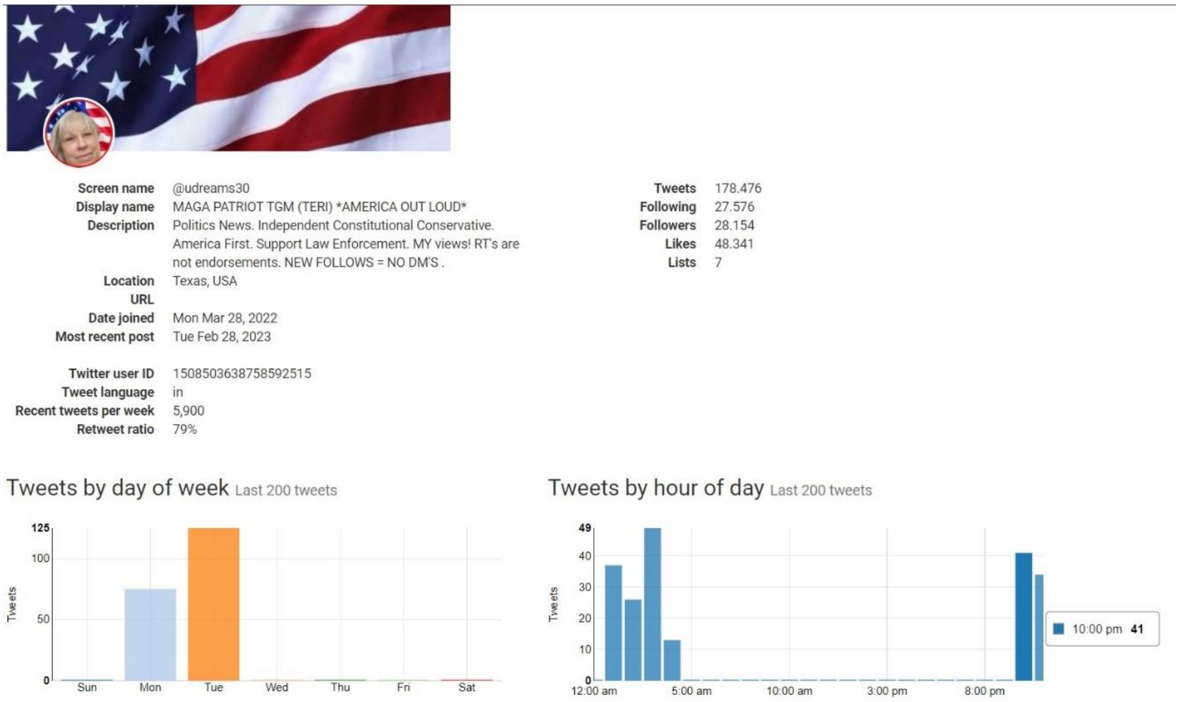
Majority tweet language: en

@thescheier 0.5 / 5

@Richard11526 4 / 5

Figure 5 - Botometer® algorithmic evaluation

Greater scores (from 0 to 5) indicate increased bot-like activities. This service requires Twitter login and authorization in order to use.



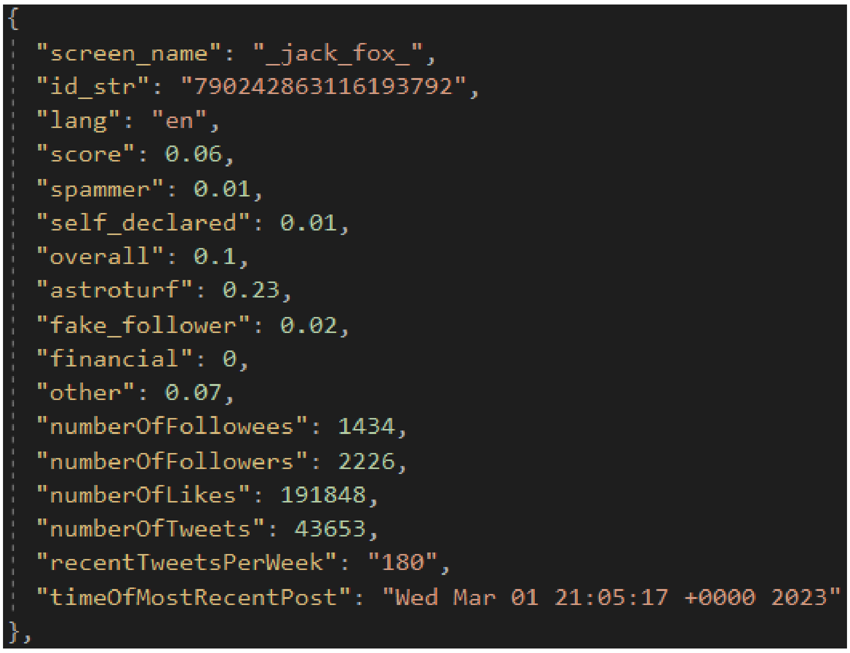
Screen name	@udreams30
Display name	MAGA PATRIOT TGM (TERI) *AMERICA OUT LOUD*
Description	Politics News. Independent Constitutional Conservative. America First. Support Law Enforcement. MY views! RT's are not endorsements. NEW FOLLOWS = NO DM'S .
Location	Texas, USA
URL	
Date joined	Mon Mar 28, 2022
Most recent post	Tue Feb 28, 2023
Twitter user ID	1508503638758592515
Tweet language	in
Recent tweets per week	5,900
Retweet ratio	79%

Tweets	178.476
Following	27.576
Followers	28.154
Likes	48.341
Lists	7

Figure 6 - Botometer® evaluation in details

However, to analyse an account, a user's nickname is needed, which is a fundamental stage of all work. By using the Twitter’s advanced search engine, it's needed to enter keywords/hashtags of specified events and select the time period for searching for tweets that will be taken as the basis of the study.

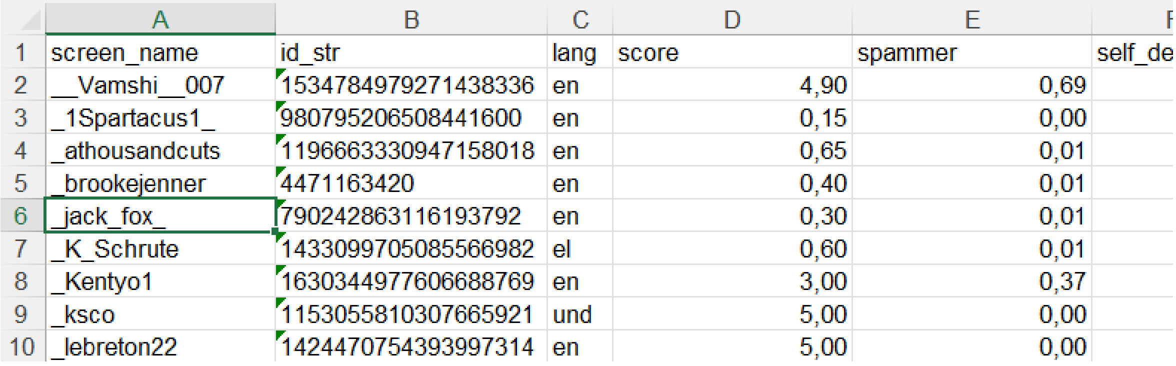
Further, from the received data array, it was necessary to select the mentioned above nicknames of the authors of the posts and only then enter the request in Botometer®. For a broader collection of data, the author decided to stop at 3,500 tweets in the 10-month span for each keyword/hashtag, which equates to 350 tweets per month or (on average) 12 tweets per day. Considering possible errors, it was decided to choose random 15 users per day.



```
{
  "screen_name": "_jack_fox_",
  "id_str": "790242863116193792",
  "lang": "en",
  "score": 0.06,
  "spammer": 0.01,
  "self_declared": 0.01,
  "overall": 0.1,
  "astroturf": 0.23,
  "fake_follower": 0.02,
  "financial": 0,
  "other": 0.07,
  "numberOfFollowees": 1434,
  "numberOfFollowers": 2226,
  "numberOfLikes": 191848,
  "numberOfTweets": 43653,
  "recentTweetsPerWeek": "180",
  "timeOfMostRecentPost": "Wed Mar 01 21:05:17 +0000 2023"
},
```

Figure 7 - Botometer® output in JSON

The final output from the Botometer® algorithm was provided as a JSON array, which, for convenience in further use, had to be converted to an Excel table version. It is worth noting, that although the algorithm initially calculated bot-like activity of given accounts in the range from 0 to 5, as a result, the value of the output was compressed to a mark from 0 to 1. This is solved by multiplying score column by * 5, thereby restoring the preliminary picture.



	A	B	C	D	E	F
1	screen_name	id_str	lang	score	spammer	self_de
2	__Vamshi__007	1534784979271438336	en	4,90	0,69	
3	_1Spartacus1_	980795206508441600	en	0,15	0,00	
4	_athousandcuts	1196663330947158018	en	0,65	0,01	
5	_brookejenner	4471163420	en	0,40	0,01	
6	_jack_fox_	790242863116193792	en	0,30	0,01	
7	_K_Schrute	1433099705085566982	el	0,60	0,01	
8	_Kentyo1	1630344977606688769	en	3,00	0,37	
9	_ksco	1153055810307665921	und	5,00	0,00	
10	_lebreton22	1424470754393997314	en	5,00	0,00	

Figure 8 - Converted JSON in spreadsheet format

4.3 Analysis Methods

4.3.1 Statistical Analysis

As a statistical analysis tool, SAS® OnDemand for Academics: Studio (online version) was employed by the author of this paper. In order to address the previously stated objectives, that formed the basis for this study, a number of diagnostic quantitative tests have been performed.

4.3.2 Descriptive Statistics

Descriptive statistics is an important tool for summarizing and analysing data in many research fields. According to Field (2013), descriptive statistics involve methods for organizing, summarizing, and presenting data using measures such as frequency tables, graphs, and summary statistics. Descriptive statistics are often used to provide a preliminary overview of data before further analysis is conducted. For example, in a study of the effects of a new teaching method on student learning, descriptive statistics may be used to summarize the distribution of student scores on a pre-test before the intervention is implemented.

Some of the key descriptive statistics calculated in this research include measures of central tendency such as the mean, median, and mode, and measures of dispersion such as the range, standard deviation, and variance. These measures can help to provide a clear picture of the distribution of the data and can assist in identifying any outliers or anomalies that may be present. Additionally, descriptive statistics can provide insight into the degree of variability within the data and can be used to compare different groups or samples, which is in this case related to particular keywords. Overall, the calculation of appropriate descriptive statistics is critical for any data analysis as it helps to provide a clearer understanding of the data and can inform further analysis and interpretation.

4.3.3 Correlation Analysis

Correlation analysis is a statistical technique used to investigate the relationship between two or more variables. According to Field (2013), correlation analysis involves calculating a correlation coefficient that measures the strength and direction of the relationship between variables. The correlation coefficient ranges from -1 to +1, with values closer to -1 indicating a strong negative correlation, values closer to +1 indicating a strong positive correlation, and values close to 0 indicating no correlation. Correlation analysis can be useful for identifying patterns and associations between variables in many research fields.

Assuming that there is a relationship between the spammer/fake follower score(s) and the account bot scores, it is crucial to clarify correlation between these variables. By calculating a correlation coefficient and considering the direction and strength of the relationship, it is possible to gain insights into patterns and associations within applied data. The strength of the relationship involving fake followers, followers, spammer, number of tweets, and score will be assessed in this research employing Pearson's correlation.

4.3.4 N-Way ANOVA

N-way ANOVA, also known as factorial ANOVA, is a statistical method that is commonly used in experimental designs with multiple independent variables. The purpose of N-way ANOVA is to determine the effect of each independent variable, as well as any interaction effects between variables, on the dependent variable. This method is widely used in various fields of research, such as psychology, medicine, and engineering.

One important aspect of N-way ANOVA is the interpretation of the main effects and interaction effects. The main effects represent the unique contribution of each independent variable to the dependent variable, while the interaction effects represent the joint effects of two or more independent variables on the dependent variable. The interpretation of these effects can be further enhanced by using post-hoc tests, such as Tukey's HSD or Bonferroni correction.

N-way ANOVA is a powerful tool for analysing the effects of multiple independent variables on a dependent variable. In the context of the thesis, ANOVA was deployed to help address the research question and main objective by identifying significant differences in the mean values of disinformation spread by different networks of Twitter accounts. By comparing the means of different groups, the researcher was able to identify statistically significant differences between groups, which provided valuable insights into the behaviour of these networks and helped to guide further analysis and interpretation of the data. However, proper interpretation and consideration of assumptions are necessary for accurate results and conclusions.

5 Results and Discussion

5.1 Descriptive Statistics

The computation of descriptive statistics, which must always come before any hypothesis testing, is a crucial initial phase in conducting investigation. The descriptive statistics chapter also involves indicators of frequency, central tendency, and variance along with additional types of variables (nominal, ordinal, interval, and ratio). Descriptive statistics, especially in this work, are critical since they aid in assessing bot-like behavioural features. They present a concise summary of the findings. Table 1 below gives an overview of the descriptive statistics of the dataset used in analysis.

Event	Number of Observations	Mean	Standard Deviation	Min	Max	Median	Variance	Mode
Covid-19	3693	2.49	1.72	0	5	2.10	2.96	5.0
Election	3619	2.15	1.57	0	5	1.65	2.46	0.3
War	3675	2.26	1.65	0	5	1.75	2.71	5.0

Table 1 - Descriptive statistics. Analysis variable – score.

The low level of mean of Botometer®'s bot score – in the scale from 0 to 5, where 0 is more likely to be a human, and 5 tends to be a bot – and even lower mode (the most frequent number in the data set) of elections in comparison with the other two sources is evident. While the mean stands for relatively small bot score among the election sample, the mode clearly indicates an overwhelming recurrence rate of 0.3 bot-like activity scores in results.

It might be even better shown on the histograms below. All the tables have already been combined and additional categorical variables "source" has been added do distinguish origin of data.

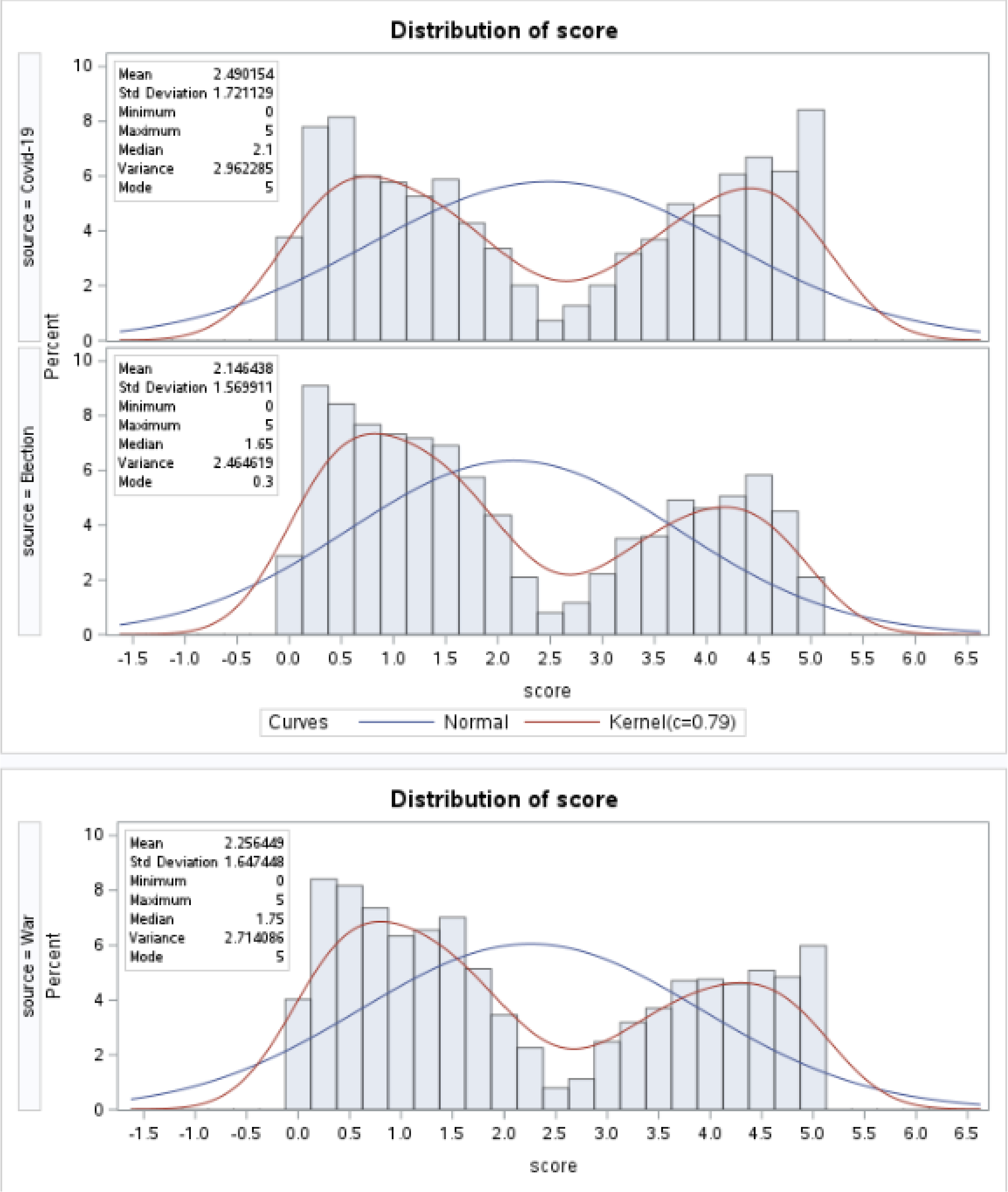


Figure 9 - Distribution of score histograms. Classification variable – source (event.)

Again, if to compare distribution of Covid-19 and Election histograms, it's easy to spot a "skew" towards the less suspicious accounts in the election sample who took part in writing Twitter posts compared to roughly the same level of the other two groups.

In order to make things clearer, author is going to add categorical variable (bot_score_cat) based on score classification in Botometer®, as if score is less than or equal to 1, it will be categorized as 1 (less suspicious) and vice versa bot score between 4 and 5 categorized as 5 (very suspicious). Is® SAS Studio it might be done via:

```
if score <= 1 then bot_score_cat = 1;
else if 1 < score <= 2 then bot_score_cat = 2;
else if 2 < score <= 3 then bot_score_cat = 3;
else if 3 < score <= 4 then bot_score_cat = 4;
else bot_score_cat = 5;
```

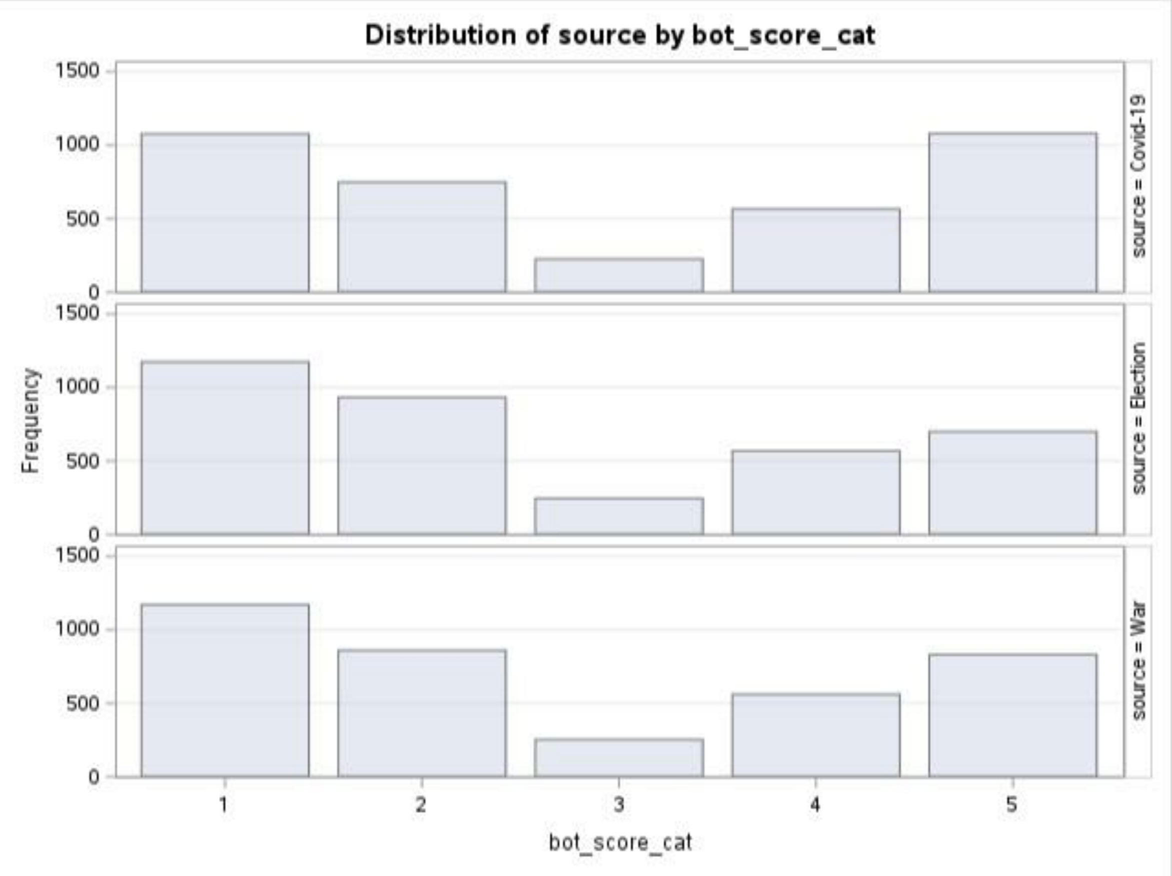


Figure 10 - Frequency Distribution of score by bot_score_cat.

It will be very beneficial for future analysis, since it is requested to combine the data obtained for the intermediate picture.

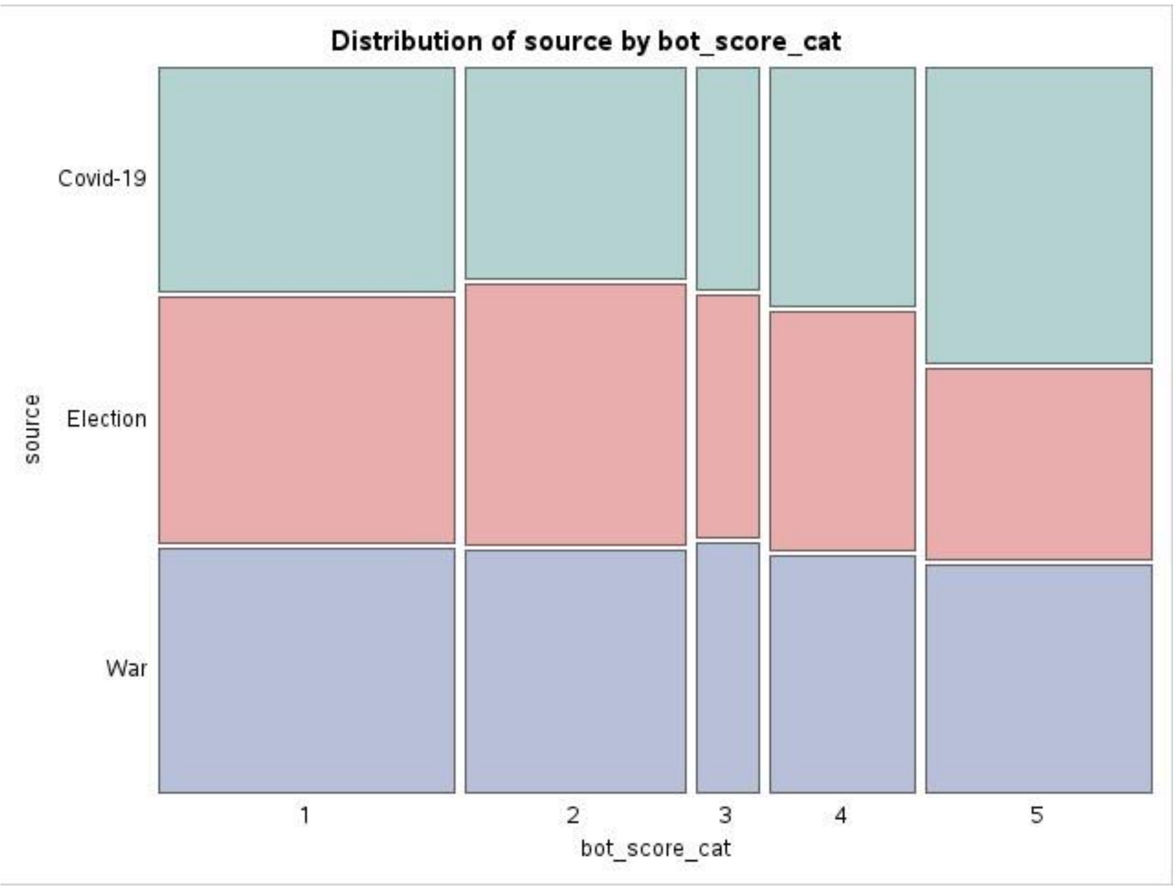


Figure 11 - Mosaic plot. Frequency Distribution relationship between events.

5.1.1 Correlation Analysis

The Pearson's correlation between the author's selection of the fake follower, spammer, numberOfFollowers, numberOfTweets, and score characteristics indicating bot behaviour is shown in Figure 12 below. A statistical estimate of the strength of the connection between the relative changes among two variables is the correlation coefficient. The range of values is from -1 to 1. Correlation coefficients less than one reflect a perfect negative relation, whereas correlation coefficients greater one implies a perfect positive relation. The absence of a normal relationship between the changes of the 2 variables is indicated by a correlation coefficient of 0. The direction and strength of this tendency to change in parallel in statistics are defined by correlation coefficients.

Variable	Total	Mean	Standard Deviation	Sum	Minimum	Maximum
score	10987	2.299	1.654	25257	0	5.0
fake_follower	10987	0.392	0.309	4303	0	1.0
spammer	10987	0.180	0.232	1982	0	1.0
numberOfFollowers	10987	46479	1184522	510667461	0	87565462
numberOfTweets	10987	15338	62278	168520692	1	2219252
Pearson Correlation Coefficient, Total (observations) = 10987						
Variable	fake_follower	spammer	numberOfFollowers		numberOfTweets	
score	0.816	0.650	-0.006		0.025	

Table 2 - Pearson Correlation

Correlation coefficient 0.8 stands for a strong positive relationship between score and fake_follower. Whereas relationship between spammer and score is not so significant, considered as moderate positive relationship. Therefore, it is enough to claim that H_1 (main) hypothesis – There is a relationship between the spammer/fake follower score(s) and the account bot scores – is confirmed.

It is worth noting that the correlation coefficient of the number of tweets and followers close to zero may characterize the lack of dependence on these indicators.

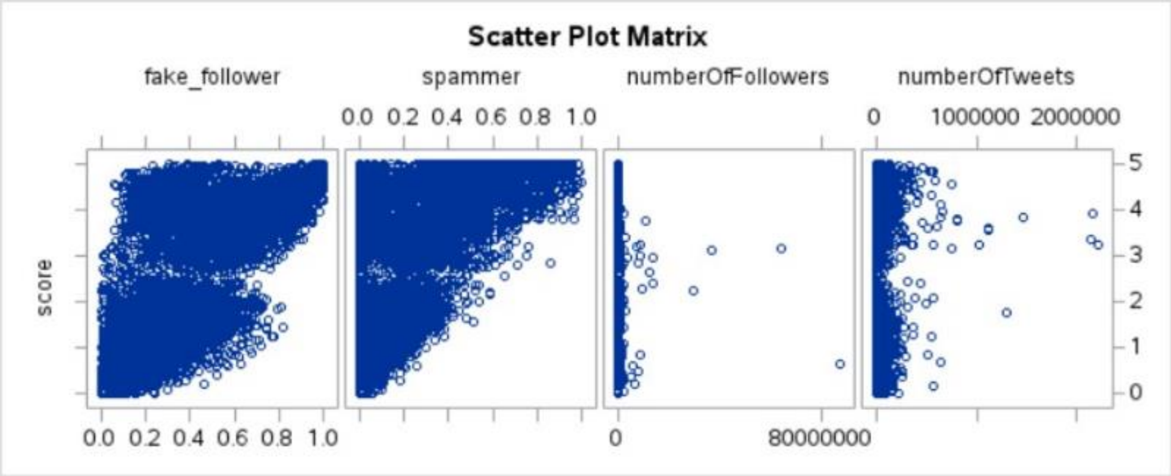


Figure 12 – Scatter Plot Matrix.

Also, a correlation analysis was conducted with the identical variables that were investigated per each event/source independently, and the results are shown in the Appendix section.

5.2 N-Way ANOVA

In N-Way ANOVA, the F-statistic measures the ratio of the variability between the groups to the variability within the groups. A larger F-value indicates a larger difference between the group means, and a smaller p-value (also represented as Pr > F) indicates that this difference is statistically significant.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	121	736.959059	6.090571	213.00	<.0001
Error	10865	310.669456	0.028594		
Corrected Total	10986	1047.628515			

Table 3 - N-Way ANOVA for Dependent variable – fake_follower

In the context of Table 2 and Table 3, an F-value of 213.00 and 75.92, respectively, and a p-values of less than 0.001 indicate that there is a significant difference between the groups being compared in the ANOVA analysis. Specifically, it suggests that the variability between the groups is much larger than the variability within the groups, and that these differences is unlikely to have occurred by chance alone.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	121	271.00	2.24	75.92	<.0001
Error	10865	320.51	0.03		
Corrected Total	10986	591.52			

Table 4 - N-Way ANOVA for Dependent variable – spammer

Therefore, it is possible to conclude that there is a significant effect of the independent variables – score and source – on the dependent variable – fake_follower/spammer – being studied. Alternative (H_1) hypothesis regarding correlation between the spammer/fake follower score(s) and the account bot scores – confirmed, in this matter. All the additional related plots might be found in Appendix section.

5.2.1 Independent vs Dependent Variable relationship

Tables 4 and 5 below represent data of mean bot score values and mean fake followers/spammers scores for each category relative to certain events.

Event/Score	Bot_Score_Cat				
	1	2	3	4	5
Covid-19	0.48	1.49	2.50	3.59	4.61
Elections	0.51	1.49	2.47	3.58	4.50
War	0.49	1.48	2.48	3.58	4.59
	Fake_follower				
Covid-19	0.12	0.29	0.39	0.59	0.83
Election	0.12	0.27	0.34	0.48	0.56
War	0.12	0.29	0.39	0.59	0.81

Table 5 - Relationship of means between fake followers and score.

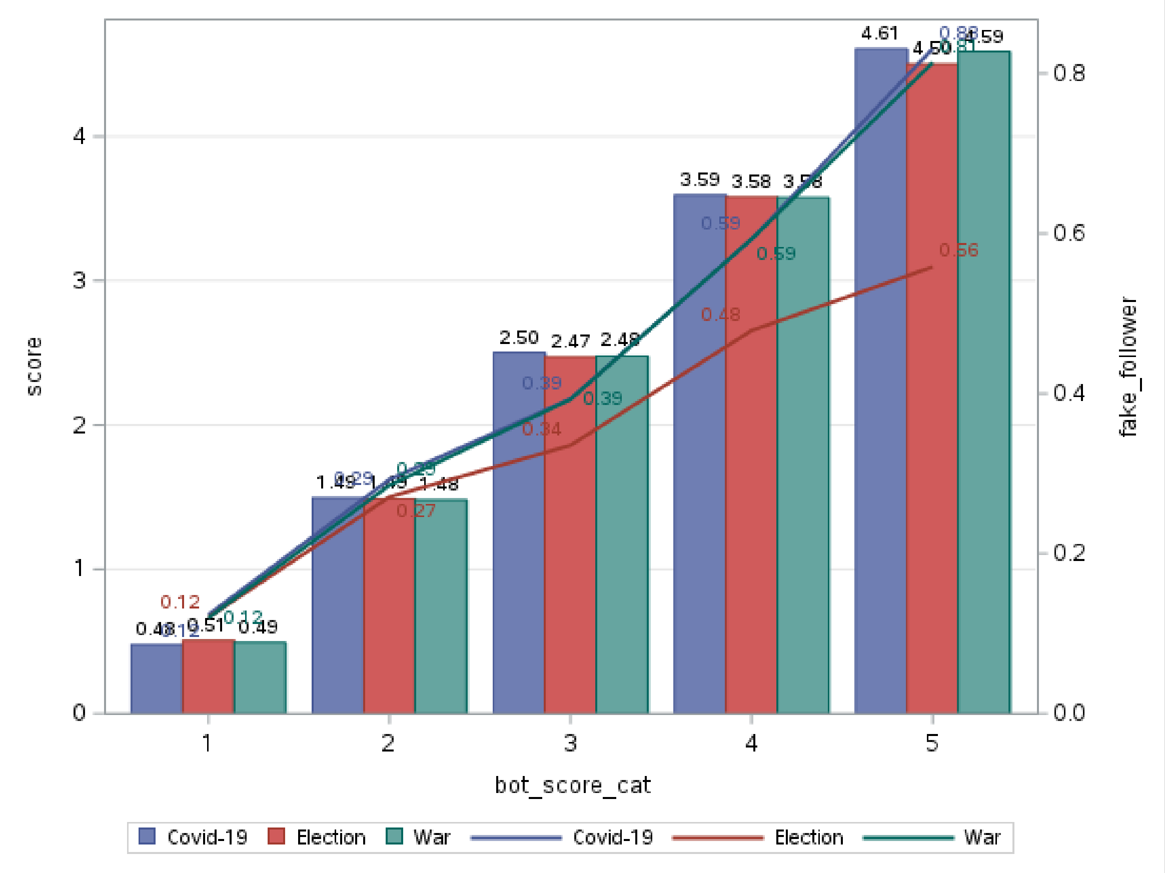


Figure 13 - Relationship of means between fake followers and score.

Based on this data, it is possible to build Bar-Line Charts (Fig. 13 and 14) for clarity of results.

Event/Score	Bot_Score_Cat				
	1	2	3	4	5
Covid-19	0.48	1.49	2.50	3.59	4.61
Elections	0.51	1.49	2.47	3.58	4.50

War	0.49	1.48	2.48	3.58	4.59
	Spammer				
Covid-19	0.02	0.11	0.23	0.28	0.45
Election	0.02	0.09	0.18	0.22	0.32
War	0.02	0.12	0.21	0.28	0.42

Table 6 - Relationship of means between spammer and score.

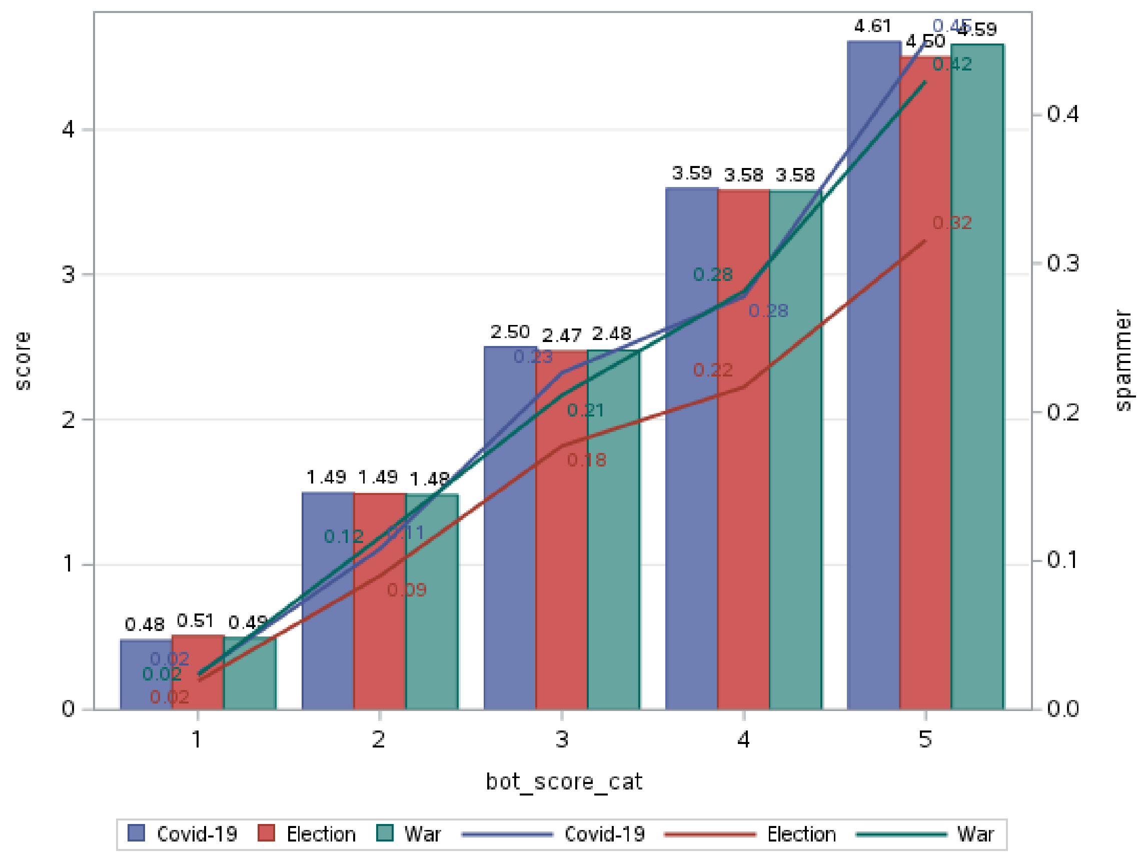


Figure 14 - Relationship of means between spammer and score.

Considering correlation analysis result and distribution frequency from previous chapters, in addition to the mean of dependent variables (fake follower / spammer) in Figures 13 and 14, this phenomenon can be characterized as – among the randomly selected 3619 tweets of users, the vast majority are real people interested in the topic of elections. The low spammer level might be even considered as more typical to bot-like action.

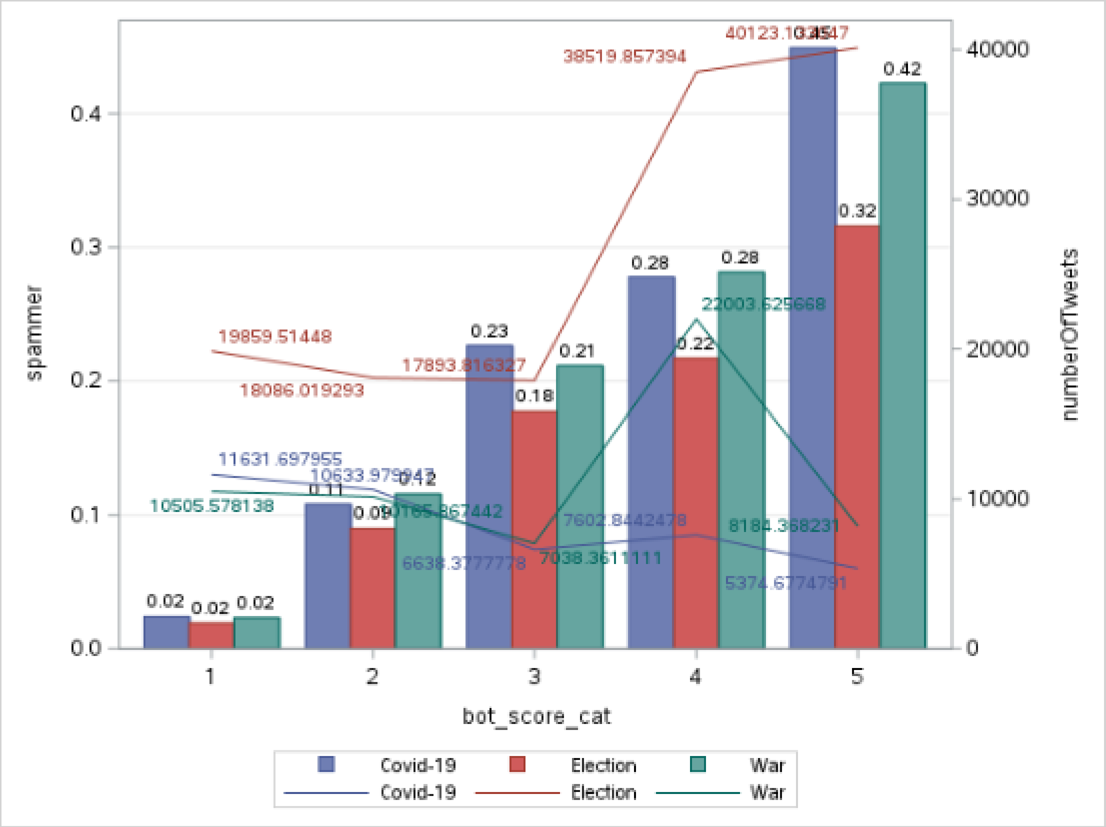


Figure 15 - Relationship of spammer to number of tweets per score category.

Bars in Figure 15 shows mean of spammer score, while lines stand for mean number of tweets. Thus, the sample size of election users has the highest number of tweets with lowest number of spammer value among bot score from 4 to 5. While the highest posting rate of ‘Covid-19’ (5374) and ‘War’ (8184) group users have almost 5–7.5 times less posts in comparison with ‘Election’ (40123) group. Therefore, number of tweets is meaningful dataset with negative relation to bot score, that helps to distinguish low active bots and users with a high activity in the network.

All other things being equal, the tendency of dependence on the bot score remains unchanged.

5.3 Discussions

In the most recent research Yang et al. (2022) collected 2000 tweets of two cryptocurrency ‘cashtags’ ($FLOKI and $SHIB) and the ‘cashtag’ of Apple Inc. ($AAPL) to quantify which is more amplified by bot-like accounts. The research considers the accounts with scores higher than a threshold as likely bots and undertaken by using same Botometer® approach.

As a threshold, researchers took a value of 0.5 and 0.7, that are respectively equivalent to 2.5 and 3.5 from this thesis.

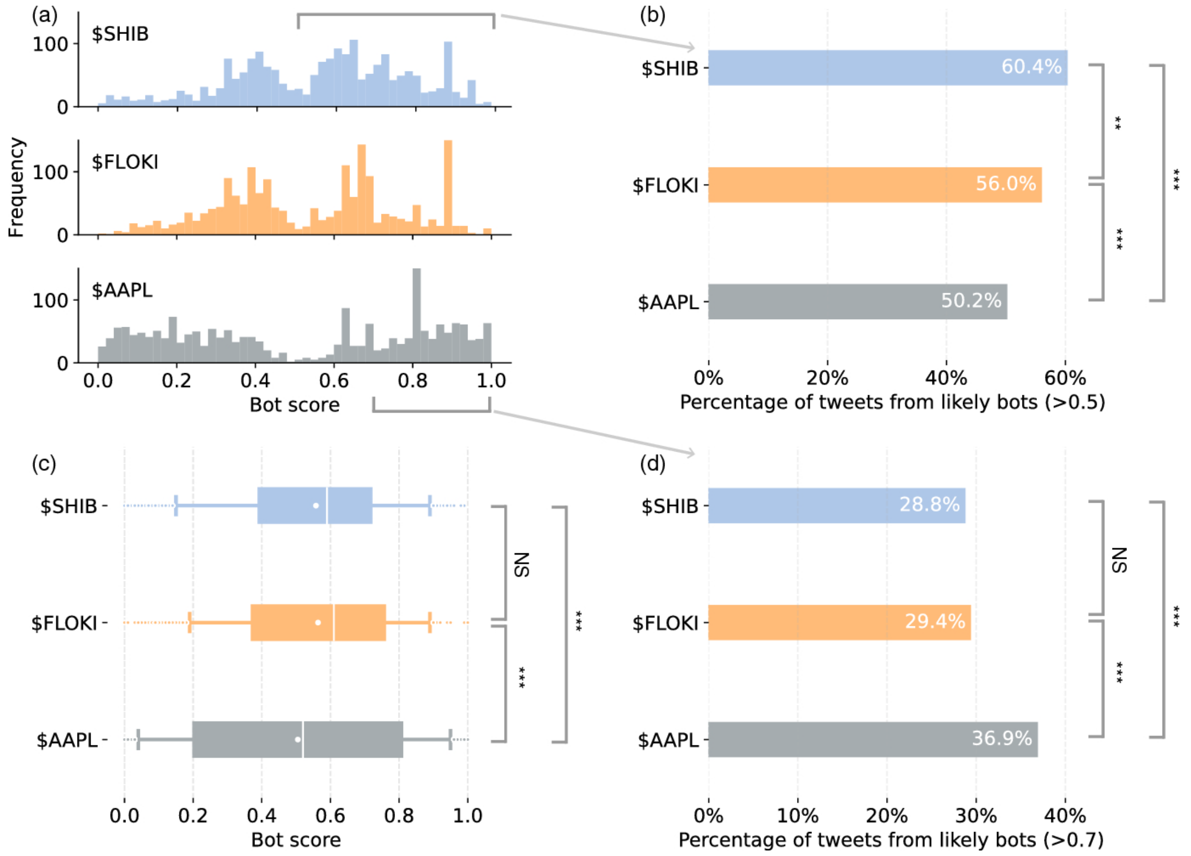


Figure 16 - Botometer 101: social bot practicum for computational social scientists (Yang et al., 2022)

Applying the same threshold Twitter bot accounts related study gives us the following results:

Score > 2.5: 47% of Covid-19, 38 % of election(s), 41 % of war related and

Score > 3.5: 37% of Covid-19, 27 % of election(s), 30 % of war related share of tweets from suspicious accounts.

The huge difference in indicators from the 0.5 threshold, according to the author, is the great interest of genuine accounts in communicating and using certain hashtags – which is definitely not in favour of cryptocurrencies. Indeed, topics cannot be matched for comparison, but the overall (> 0.7) trend remains roughly the same.

5.4 Limitations and implications

The main limitation of this work is a rather modest size of dataset. In projects of this scale, a database of several billion is needed to fully disclose the topic and get answers to the questions posed. Larger amounts of data will make it much more accurate to define the relationships of certain variables. In terms of correlation analysis, this allows to determine the dependence of independent variables regarding the bot score. Thus, beneficial in establishing a connection with new possible bot activity indicators.

The theoretical value of this work is the confirmed relationship between the bot score and the number of fake subscribers/spammers. By confirming this relationship, the research can provide valuable insights into how to detect and prevent fraudulent behaviour on various platforms that rely on subscriber or user data. This information can be particularly useful for individuals and organizations looking to mitigate the impact of fake accounts and improve the accuracy and reliability of their data.

This study could have several practical implications. For example, if a Twitter network uses a bot detection system based on bot scores, the research findings can help improve the accuracy of the system by identifying thresholds for bot scores that are associated with a high number of fake subscribers or spammer score. This can help the platform to identify and remove fake accounts more effectively, which can improve the platform's overall user experience and increase user trust. Additionally, the findings could be used to develop better tools and techniques for detecting and preventing fraudulent behaviour on various online platforms.

6 Conclusion

Main aim of this research was to identify networks of accounts that spread disinformation on Twitter. The author used primary data as a basis of study, by collecting dataset in timeframe of 10 months regarding different keywords/hashtags of specific events; elections (May 2016 – March 2017, 3619 participants), Covid-19 (March 2020 – January 2021, 3693 participants) and war (February 2022 – December 2022, 3675 participants). According to Table 6, 2069 users or 23.74% among 10987 total sample size were suspected with a bot-like behaviour. SAS® OnDemand for Academics: Studio, a statistical software, was used to examine the data. The following results were reached after assessing the study's partial objectives:

1) to make an overview of the current state of the play in identifying Twitter accounts spreading disinformation and methods of detecting them

The spread of disinformation on Twitter has become a significant concern in recent years. Many individuals and groups use the platform to disseminate false information, propaganda, and other forms of manipulative content to achieve their agendas. To tackle this issue, researchers and experts have developed various methods for identifying and detecting Twitter accounts that spread disinformation.

One of the most common methods is network analysis, which involves identifying patterns and relationships between accounts that share disinformation. Researchers also use natural language processing techniques to analyse the content of tweets and identify patterns that suggest disinformation. Machine learning algorithms are also employed to analyse large volumes of Twitter data and identify patterns that suggest the presence of disinformation. Overall, while the task of identifying Twitter accounts spreading disinformation remains challenging, ongoing research efforts are providing promising methods for detecting and mitigating this issue.

Whereas the differences in the relationship between these variables indicated that more than 90% of the individuals who tweeted about the elections were real people, 23.74% of accounts from randomly selected 10,987 users overall with a fairly high amount of certainty are classified as bots.

2) to select a computational method for bots' identification and prepare a dataset of *tweets concerning a selected topic*

Even at the stage of preparation for writing this thesis, the author came across several studies from other researchers where they used one computational algorithm – the Botometer®. The numerous choices of this method showed a credit of trust among researchers, which simplified the vector of further work.

The next key factor was the presence of a database on the basis of which the above-mentioned algorithm would work. For the purity of the experiment, the author of the study decided to create unique database based on the search for keywords of certain topic in Twitter posts.

The foundation of entire investigation was obtained by merging of these two major stages.

3) to run an experiment to detect accounts spreading disinformation

The preliminary analysis of the data indicated that over 20% of the participants were categorized as bots. This finding suggests that a significant number of accounts in the sample may not represent real people. It is important to further investigate the characteristics and behaviour of these accounts to determine the potential impact on the study's results.

The results indicate that the group of users who tweeted about the election had the highest number of tweets, but the lowest spammer value among accounts with a bot score ranging from 4 to 5. On the other hand, the groups of users who tweeted about Covid-19 and War had significantly fewer posts, ranging from 5 to 7.5 times less than the Election group. Hence, the number of tweets is a valuable variable with a negative relation to the bot score, which can help distinguish low-activity bots from highly active users in the network.

7 References

1. CALDARELLI, Guido, et al. The role of bot squads in the political propaganda on Twitter. *Communications Physics*, 2020, 3.1: 1-15.

2. MOZUR, Paul, LEE MYERS, Steven, KAO, Jeff and THE NEW YORK TIMES. How bots and fake accounts push China's vision of Winter Olympic Wonderland. *How Bots and Fake Accounts Push China's Vision of Winter Olympic Wonderland* [online]. 18 February 2022. [Accessed 5 May 2022]. Available from: https://www.propublica.org/article/how-bots-and-fake-accounts-push-chinas-vision-of-winter-olympic-wonderland

3. GRUZD, Anatoliy; MAI, Philip. Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. *Big Data & Society*, 2020, 7.2: 2053951720938405.

4. GUARINO, Stefano, et al. Characterizing networks of propaganda on twitter: a case study. *Applied Network Science*, 2020, 5.1: 1-22.

5. BELLOVARY, Andrea; YOUNG, Nathaniel A.; GOLDENBERG, Amit. Left-and Right-Leaning News Organizations' Negative Tweets are More Likely to be Shared. 2021.

6. WANG, Di; LU, Jiahui. How news agencies' Twitter posts on COVID-19 vaccines attract audiences' Twitter engagement: A content analysis. *International Journal of Environmental Research and Public Health*, 2022, 19.5: 2716.

7. JOWETT, Garth S. and O'DONNELL, Victoria. Propaganda and Persuasion - Fifth Edition. *PropagandaPersuasion2012* [online]. 2016. [Accessed 25 July 2022]. Available from: https://hiddenhistorycenter.org/wp-content/uploads/2016/10/PropagandaPersuasion2012.pdf

8. RATKIEWICZ, Jacob, et al. Detecting and tracking political abuse in social media. In: *Proceedings of the International AAAI Conference on Web and social media*. 2011. p. 297-304.

9. JOHANSEN, Alison Grace. What's a Twitter bot and how to spot one. *What's a Twitter bot and how to spot one* [online]. [Accessed 2 August 2022]. Available from: https://au.norton.com/internetsecurity-emerging-threats-what-are-twitter-bots-and-how-to-spot-them.html

10. CHU, Zi, et al. Detecting automation of twitter accounts: Are you a human, bot, or cyborg?. *IEEE Transactions on dependable and secure computing*, 2012, 9.6: 811-824.

11. VAROL, Onur, et al. Online human-bot interactions: Detection, estimation, and characterization. In: *Proceedings of the international AAAI conference on web and social media*. 2017. p. 280-289.

12. NISOS. Influencing the narrative: Nisos investigates an inauthentic Xinjiang twitter network. *Influencing the Narrative: Nisos Investigates an Inauthentic Xinjiang Twitter Network* [online]. 12 May 2022. [Accessed 17 August 2022]. Available from: https://www.nisos.com/blog/xinjiang-twitter-network-report/

13. Report: Fake twitter accounts spread Chinese propaganda. *AP NEWS* [online]. 25 April 2022. [Accessed 10 August 2022]. Available from: https://apnews.com/article/technology-business-china-beijing-race-and-ethnicity-14fec4421be0291e5f0ea580ecbd4b6d

14. SCHLIEBS, Marcel, et al. China's public diplomacy operations: understanding engagement and inauthentic amplifications of PRC diplomats on Facebook and Twitter. 2021.

15. KELLER, Franziska B., et al. Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Political Communication*, 2020, 37.2: 256-280.

16. Political figures with the most Twitter followers. *Daily Sabah* [online]. 16 February 2022. [Accessed 20 August 2022]. Available from: https://www.dailysabah.com/gallery/political-figures-with-the-most-twitter-followers/images

17. LEWIS, Peter H. And the spoof begat a news release, and another. *The New York Times* [online]. 31 December 1994. [Accessed 19 August 2022]. Available from: https://www.nytimes.com/1994/12/31/business/and-the-spoof-begat-a-news-release-and-another.html

18. RYAN, Fergus, BOGLE, Ariel, ZHANG, Albert and WALLIS, Jacob. #StopXinjiang rumors | australian strategic policy institute | ASPI. *#StopXinjiang Rumors* [online]. 2 December 2021. [Accessed 14 August 2022]. Available from: https://www.aspi.org.au/report/stop-xinjiang-rumors

19. LUKITO, Josephine, et al. The wolves in sheep’s clothing: How Russia’s Internet Research Agency tweets appeared in US news as vox populi. *The International Journal of Press/Politics*, 2020, 25.2: 196-216.

20. GRAHAM, Timothy and THOMPSON, Jay Daniel. Russian government accounts are using a Twitter loophole to spread disinformation. *The Conversation* [online]. 15 March

2022. [Accessed 16 August 2022]. Available from: https://theconversation.com/russian-government-accounts-are-using-a-twitter-loophole-to-spread-disinformation-178001

21. An army of suspicious accounts began churning out pro-Russian content in March. *Russia is swaying Twitter users outside the West to its side* [online]. 14 May 2022. [Accessed 16 August 2022]. Available from: https://www.economist.com/graphic-detail/2022/05/14/russia-is-swaying-twitter-users-outside-the-west-to-its-side

22. PRINCEWILL, Nimi. Africa's highest peak gets fast internet. *CNN* [online]. 18 August 2022. [Accessed 26 August 2022]. Available from: https://edition.cnn.com/travel/article/kilimanjaro-gets-internet-service-intl/index.html

23. GOUNARI, Panayota. Twitter Blog Post on the Permanent Suspension of Donald Trump's Account, January 8th, 2021. In: *From Twitter to Capitol Hill*. Brill, 2021. p. 162-164.

24. BELLI, Luca. *Examining algorithmic amplification of political content on Twitter* [online]. 21 October 2021. [Accessed 14 August 2022]. Available from: https://blog.twitter.com/en_us/topics/company/2021/rml-politicalcontent

25. SHEPHERD, Jack. 22 essential twitter statistics you need to know in 2023. *22 Essential Twitter Statistics You Need to Know in 2023* [online]. 3 January 2023. [Accessed 14 January 2023]. Available from: https://thesocialshepherd.com/blog/twitter-statistics

26. Dealing with propaganda, misinformation and fake news - democratic schools for all - publi.coe.int. *Democratic Schools for All* [online]. [Accessed 15 August 2022]. Available from: https://www.coe.int/en/web/campaign-free-to-speak-safe-to-learn/dealing-with-propaganda-misinformation-and-fake-news#

27. KORNBLUH, Karen, WEINER, Eli and GOLDSTEIN, Adrienne. New study by Digital New Deal finds engagement with deceptive outlets higher on Facebook today than run-up to 2016 election. *Transatlantic Take* [online]. 12 October 2020. [Accessed 10 August 2022]. Available from: https://www.gmfus.org/news/new-study-digital-new-deal-finds-engagement-deceptive-outlets-higher-facebook-today-run-2016

28. MISLOVE, Alan, LEHMANN, Sune, AHN, Young-Yeol, ONNELA, Jukka-Pekka and ROSENQUIST, Niels James. *Understanding the Demographics of Twitter Users* [online]. January 2011. [Accessed 15 January 2023]. Available from: https://www.researchgate.net/profile/Jukka-Pekka-

Onnela/publication/221297994_Understanding_the_Demographics_of_Twitter_Users/links/0deec51a629b9187c9000000/Understanding-the-Demographics-of-Twitter-Users.pdf

29. FERRARA, Emilio, VAROL, Onur, DAVIS, Clayton Allen, MENCZER, Filippo and FLAMMINI, Alessandro. *The Rise of Social Bots* [online]. July 2014. [Accessed 15 January 2023]. Available from: https://www.researchgate.net/publication/264123205_The_Rise_of_Social_Bots

30. DAVIS, Clayton Allen, VAROL, Onur, FERRARA, Emilio, FLAMMINI, Alessandro and MENCZER, Filippo. BotOrNot: A System to Evaluate Social Bots. *BotOrNot | Proceedings of the 25th International Conference Companion on World Wide Web* [online]. 11 April 2016. [Accessed 15 January 2023]. Available from: https://dl.acm.org/doi/10.1145/2872518.2889302

31. CONOVER, Michael, RATKIEWICZ, Jacob, FRANCISCO, Matthew, GONÇALVES, Bruno, MENCZER, Filippo and FLAMMINI, Alessandro. *Political Polarization on Twitter* [online]. January 2011. [Accessed 15 January 2023]. Available from: https://www.researchgate.net/publication/221297916_Political_Polarization_on_Twitter

32. CHAVOSHI, Nikan, HAMOONI, Hossein and MUEEN, Abdullah. *DeBot: Twitter Bot Detection via Warped Correlation* [online]. December 2016. [Accessed 15 January 2023]. Available from: https://www.researchgate.net/publication/308021270_DeBot_Twitter_Bot_Detection_via_Warped_Correlation

33. BRISCOE, Erica, APPLING, Scott D. and HAYES, Heather. *Cues to Deception in Social Media Communications* [online]. January 2014. [Accessed 15 January 2023]. Available from: https://www.researchgate.net/publication/262256864_Cues_to_Deception_in_Social_Media_Communications

34. FERRARA, Emilio and YANG, Zeyao. *Quantifying the Effect of Sentiment on Information Diffusion in Social Media* [online]. June 2015. [Accessed 15 January 2023]. Available from: https://www.researchgate.net/publication/278969329_Quantifying_the_Effect_of_Sentiment_on_Information_Diffusion_in_Social_Media

35. FARKAS, Johan and NEUMAYER, Christina. *Disguised Propaganda from Digital to Social Media* [online]. July 2018. [Accessed 15 January 2023]. Available from:

https://www.researchgate.net/publication/326583846_Disguised_Propaganda_from_Digital_to_Social_Media

36. Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

37. Bryman, A., & Bell, E. (2015). Business research methods. Oxford University Press.

38. Hair, J. F., Wolfinbarger, M., Ortinau, D. J., & Bush, R. P. (2017). Essentials of marketing research. McGraw-Hill Education.

39. Denzin, N. K., & Lincoln, Y. S. (2017). The SAGE handbook of qualitative research. Sage publications.

40. Babbie, E. R. (2016). The practice of social research. Cengage Learning.

41. Gay, L. R., Mills, G. E., & Airasian, P. W. (2018). Educational research: Competencies for analysis and applications. Pearson.

42. Neuman, W. L. (2014). Social research methods: Qualitative and quantitative approaches. Pearson.

43. Sattari, S., & Mazidi, A. (2020). Effects of Instagram influencer type, product type, and product endorsement on advertising effectiveness. Journal of Retailing and Consumer Services, 53, 101738. https://doi.org/10.1016/j.jretconser.2019.101738

44. Field, A. (2013). Discovering statistics using IBM SPSS statistics. Sage publications.

45. Kirk, R. E. (2012). Experimental design: Procedures for the behavioral sciences. SAGE Publications.

46. Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics. Pearson.

47. CUNCIC, Arlin. How does propaganda work? *How Does Propaganda Work?* [online]. 12 April 2022. [Accessed 13 August 2022]. Available from: https://www.verywellmind.com/how-does-propaganda-work-5224974

48. Orabi, H., Al-Emran, M., & Shaalan, K. (2020). Detection of Bots in Social Media: A Systematic Review. International Journal of Information Management, 50, 96-133. doi: 10.1016/j.ijinfomgt.2019.05.019

49. Babbie, E. (2013). The basics of social research. Cengage Learning.

50. Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), The SAGE handbook of qualitative research (3rd ed., pp. 191-215). Sage Publications.

51. Popper, K. R. (1959). The logic of scientific discovery. Routledge.

52. Walliman, N. (2017). Research methods: the basics. Routledge.

53. YANG, Kai-Cheng, FERRARA, Emilio and MENCZER, Filippo. Botometer 101: Social Bot Practicum for computational social scientists - journal of computational social science. *SpringerLink* [online]. 20 August 2022. [Accessed 30 August 2022]. Available from: https://link.springer.com/article/10.1007/s42001-022-00177-5#Sec8

8 List of pictures, tables, graphs and abbreviations

8.1 List of pictures

8.2 List of tables

8.3 List of graphs

8.4 List of abbreviations

API – Application Programming Interface

CNN-LSTM – Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM)

DNA – Deoxyribonucleic Acid

JSON – JavaScript Object Notation

Appendix

Event	**Bot Score**					**Total**
	1	2	3	4	5	
Covid-19	1076	748	225	565	1079	3693
(in %)	9.79	6.81	2.05	5.14	9.82	33.61
Elections	1174	933	245	568	699	3619
(in %)	10.69	8.49	2.23	5.17	6.36	32.94
War	1171	860	252	561	831	3675
(in %)	10.66	7.83	2.29	5.11	7.56	33.45
Total	3421	2541	722	1694	2609	10987
(in %)	31.14	23.13	6.57	15.42	23.74	100

Table 7 - Complementary to Mosaic plot (Figure 11.)

Distribution Analysis – Plateau or Multimodal Distribution

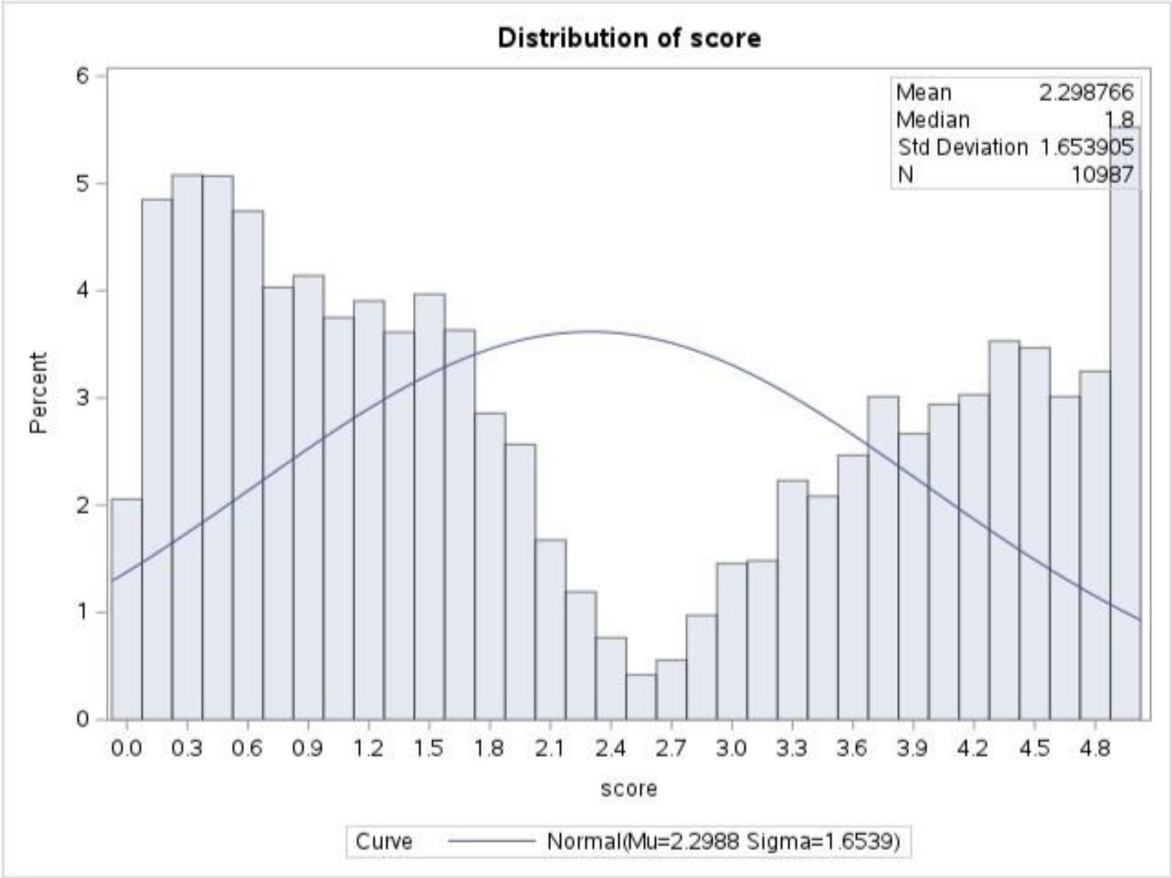
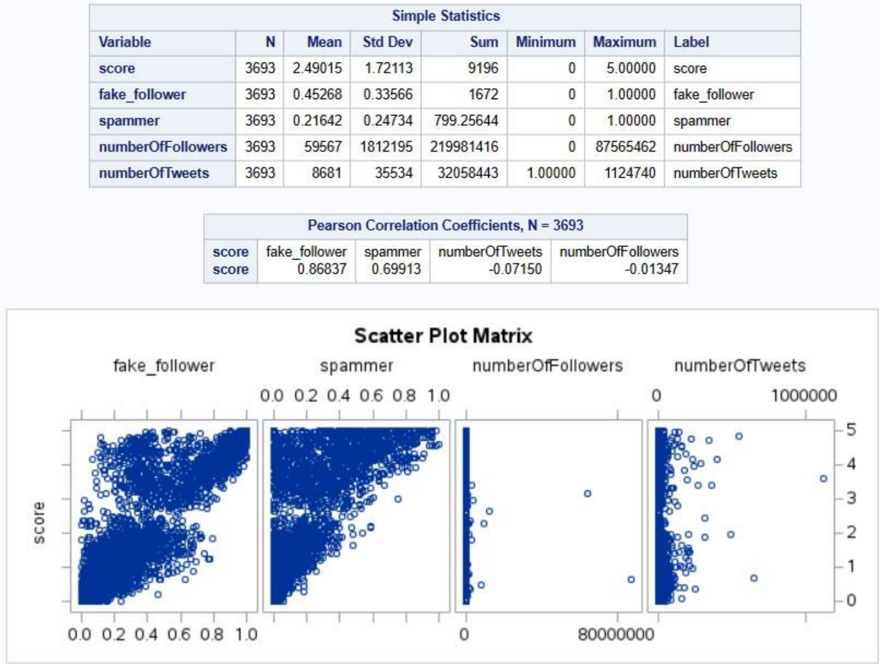


Figure 17 - Distribution Analysis – Plateau or Multimodal Distribution of score (general)

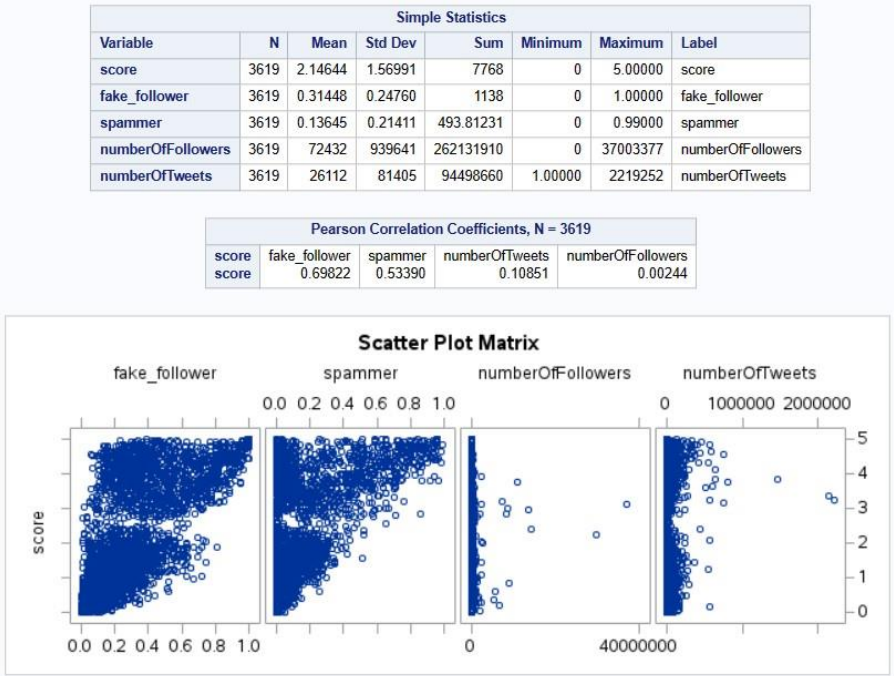
The histogram above clearly shows that the multimodal distribution indicates findings from various subpopulations, in our case – different events, and natural variations of data, primary data taken from Twitter social network.



Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
score	3693	2.49015	1.72113	9196	0	5.00000	score
fake_follower	3693	0.45268	0.33566	1672	0	1.00000	fake_follower
spammer	3693	0.21642	0.24734	799.25644	0	1.00000	spammer
numberOfFollowers	3693	59567	1812195	219981416	0	87565462	numberOfFollowers
numberOfTweets	3693	8681	35534	32058443	1.00000	1124740	numberOfTweets

Pearson Correlation Coefficients, N = 3693				
score	fake_follower	spammer	numberOfTweets	numberOfFollowers
score	0.86837	0.69913	-0.07150	-0.01347

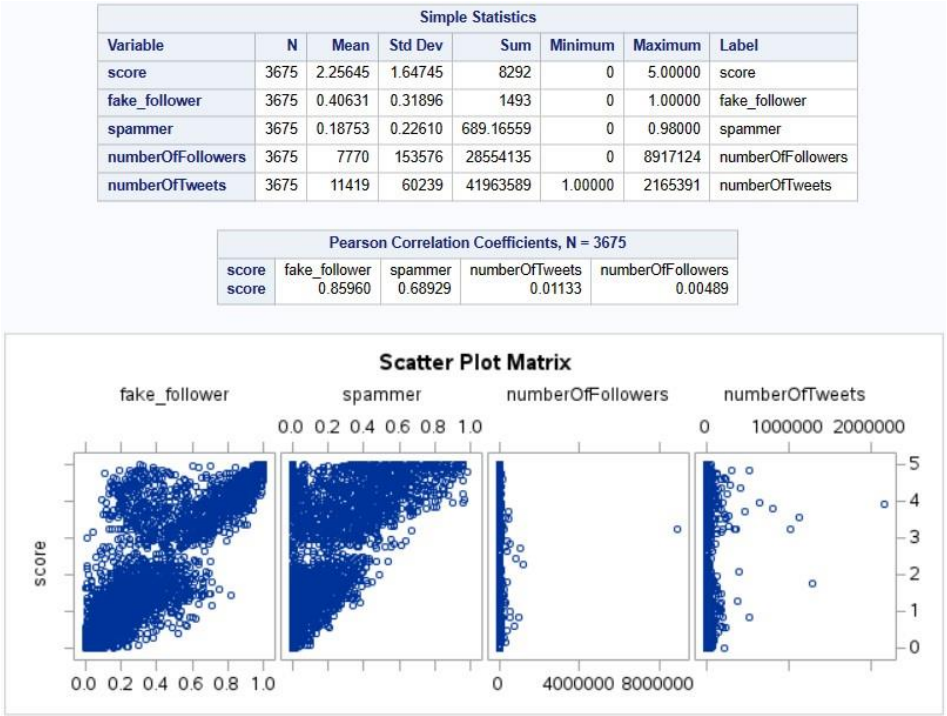
Figure 18 - Correlation analysis (Covid-19)



Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
score	3619	2.14644	1.56991	7768	0	5.00000	score
fake_follower	3619	0.31448	0.24760	1138	0	1.00000	fake_follower
spammer	3619	0.13645	0.21411	493.81231	0	0.99000	spammer
numberOfFollowers	3619	72432	939641	262131910	0	37003377	numberOfFollowers
numberOfTweets	3619	26112	81405	94498660	1.00000	2219252	numberOfTweets

Pearson Correlation Coefficients, N = 3619				
score	fake_follower	spammer	numberOfTweets	numberOfFollowers
score	0.69822	0.53390	0.10851	0.00244

Figure 19 - Correlation analysis (elections)



Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
score	3675	2.25645	1.64745	8292	0	5.00000	score
fake_follower	3675	0.40631	0.31896	1493	0	1.00000	fake_follower
spammer	3675	0.18753	0.22610	689.16559	0	0.98000	spammer
numberOfFollowers	3675	7770	153576	28554135	0	8917124	numberOfFollowers
numberOfTweets	3675	11419	60239	41963589	1.00000	2165391	numberOfTweets

Pearson Correlation Coefficients, N = 3675				
	fake_follower	spammer	numberOfTweets	numberOfFollowers
score score	0.85960	0.68929	0.01133	0.00489

Figure 20 - Correlation analysis (war)

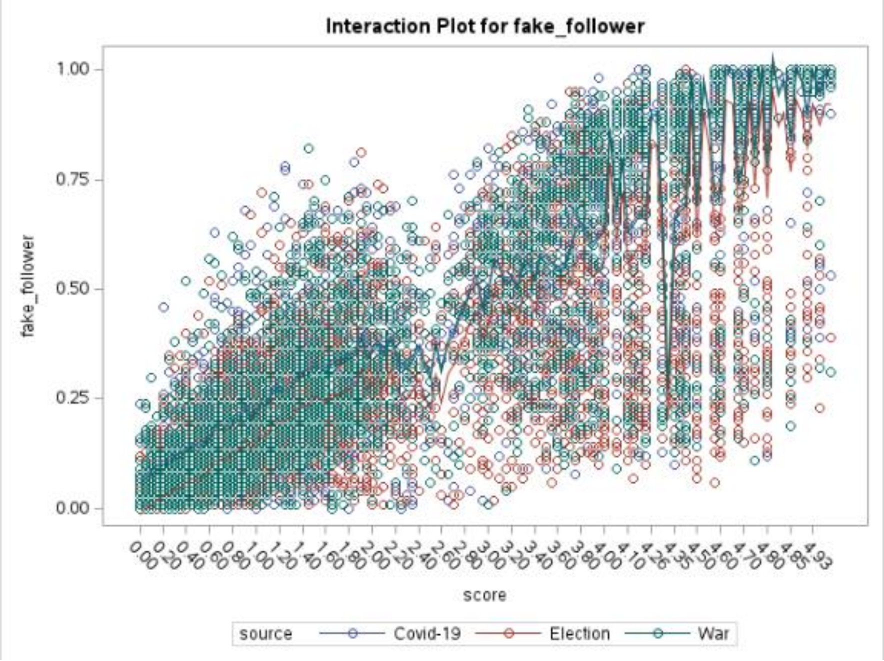


Figure 21 - Interaction Plot for (dependent variable) fake_follower. Factors – score, source.

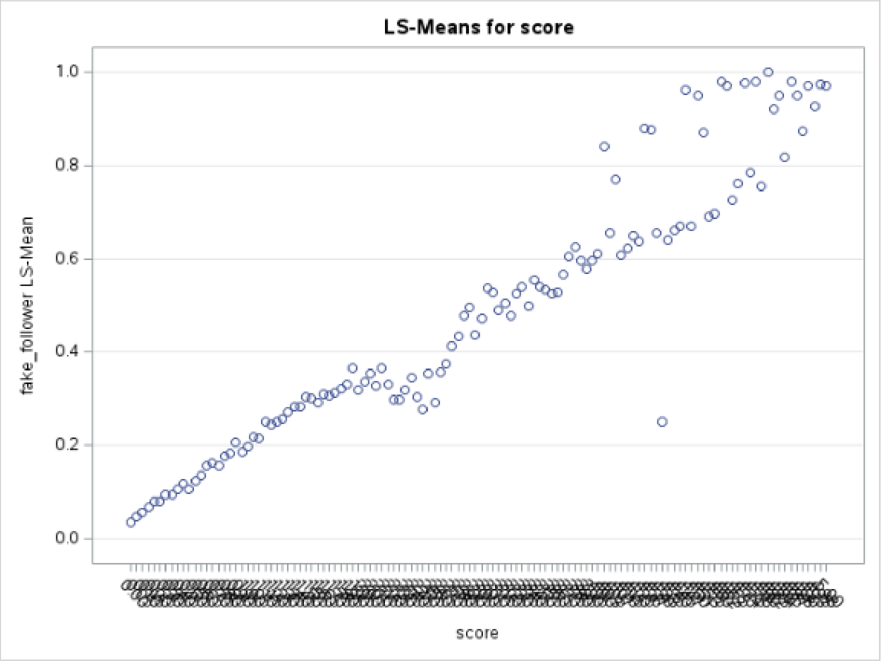


Figure 22 - LS-Means for score. Dependent variable - Fake_follower

Event	Least Squares Means	95% Confidence Limits	
Covid-19	0.514	0.502	0.525
Election	0.436	0.424	0.447
War	0.505	0.493	0.516

Table 8 - Adjustment for Multiple Comparison: Tukey-Kramer. Dependent variable – fake_follower

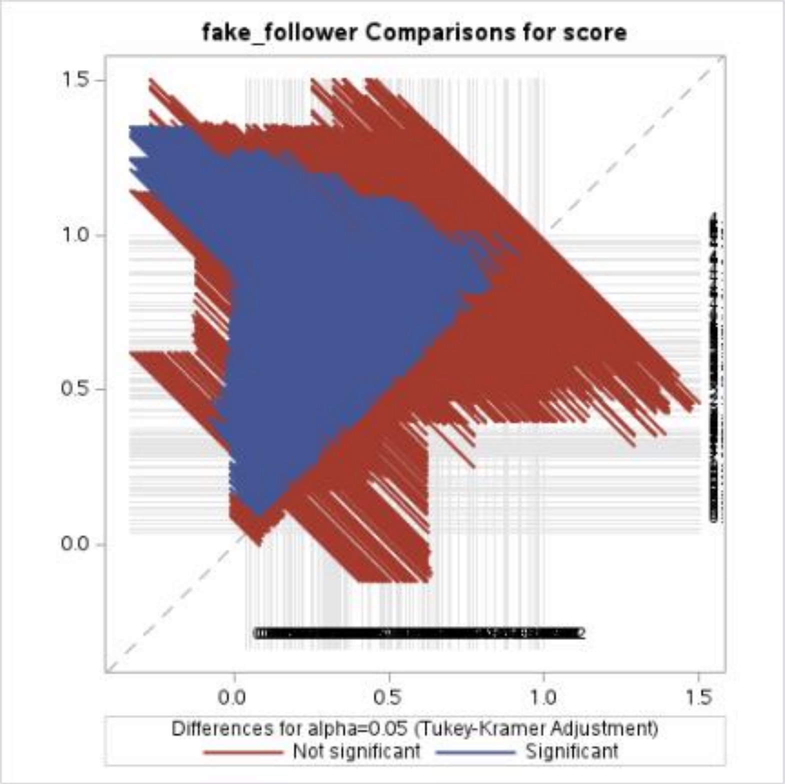


Figure 23 - Turkey-Kramer Adjustment for fake_follower (general)

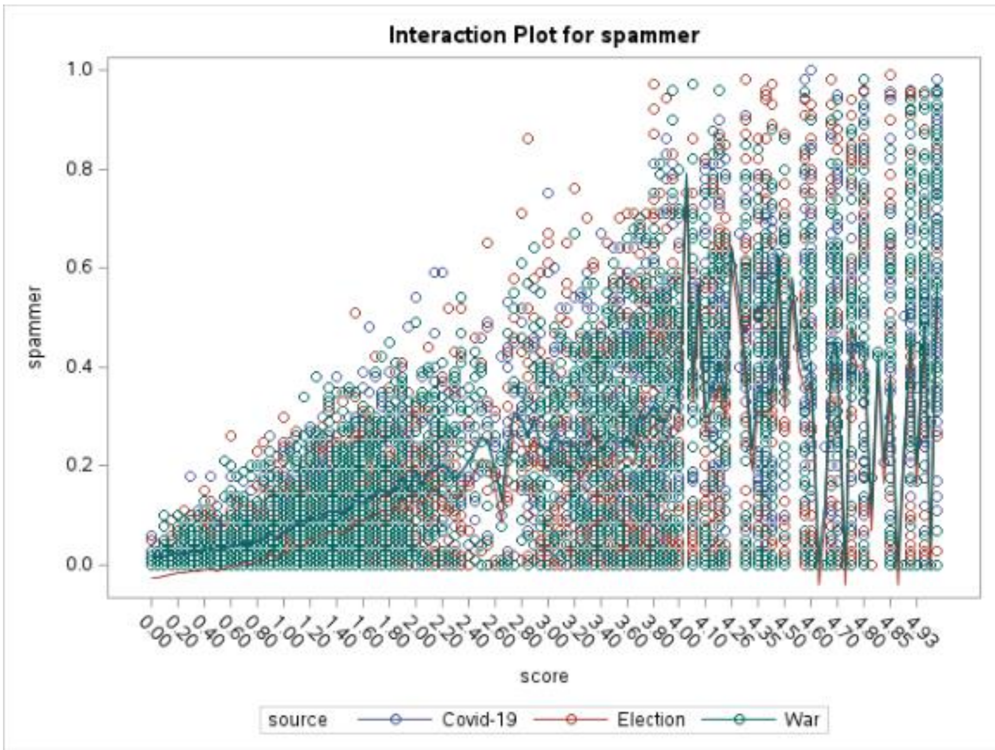


Figure 24 - Interaction Plot for (dependent variable) spammer. Factors – score, source.

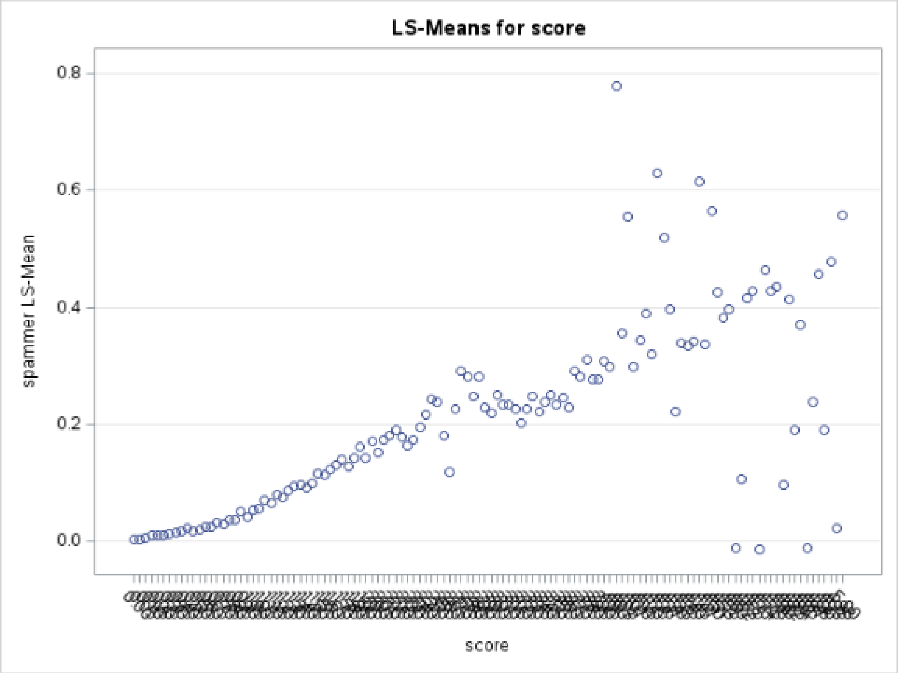


Figure 25 - LS-Means for score. Dependent variable - spammer

Event	Least Squares Means	95% Confidence Limits	
Covid-19	0.230	0.218	0.242
Election	0.187	0.175	0.199
War	0.227	0.215	0.238

Table 9 - Adjustment for Multiple Comparison: Tukey-Kramer. Dependent variable - spammer.

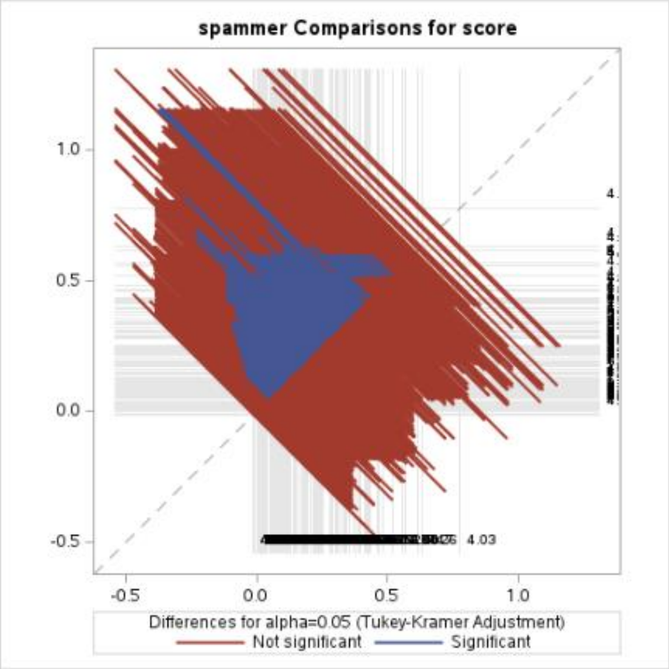


Figure 26 - Turkey-Kramer Adjustment for spammer (general)

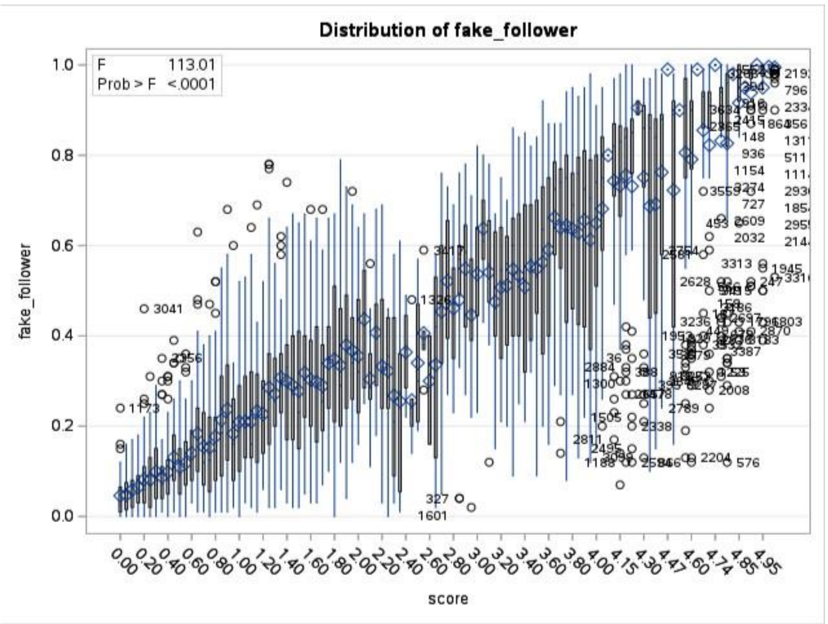


Figure 27 - distribution of fake follower (Covid-19)

Bot Score	Data Min.	Min. Whisker	1st Quartile	Median	3rd Quartile	Max. Whisker	Data Max.	Mean	Standard Deviation	Number of Observations
1	0.00	0.00	0.05	0.10	0.17	0.35	0.68	0.12	0.10	1076
2	0.00	0.00	0.18	0.27	0.39	0.69	0.79	0.29	0.16	748
3	0.00	0.00	0.26	0.41	0.54	0.82	0.82	0.39	0.18	225
4	0.08	0.08	0.42	0.65	0.76	0.98	0.98	0.59	0.20	565
5	0.07	0.57	0.82	0.91	0.99	1.00	1.00	0.83	0.22	1079

Table 10 - Box plot basis - fake_follower (Covid-19)

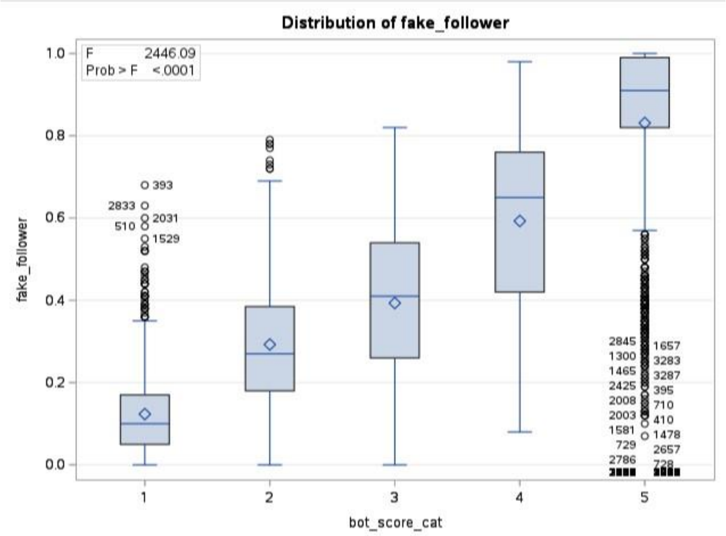


Figure 28 - Box plot - fake_follower (Covid-19)

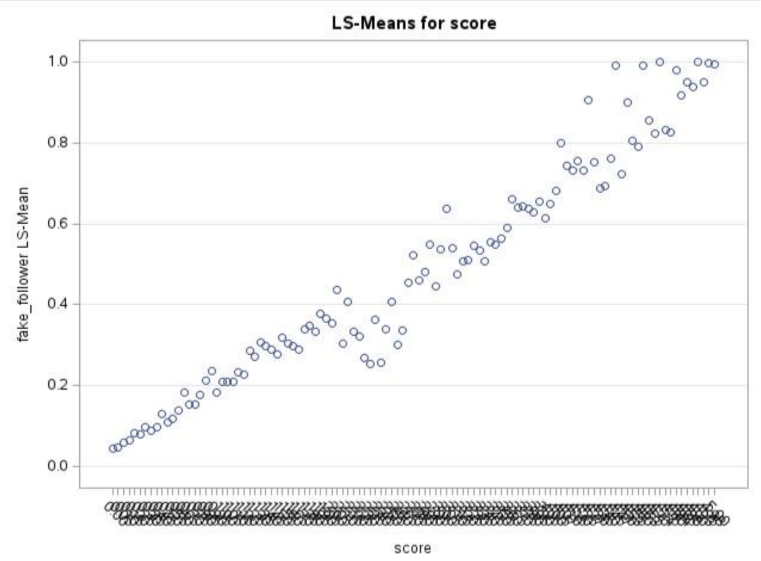


Figure 29 - Fake_follower Least Square Means for score (Covid-19)

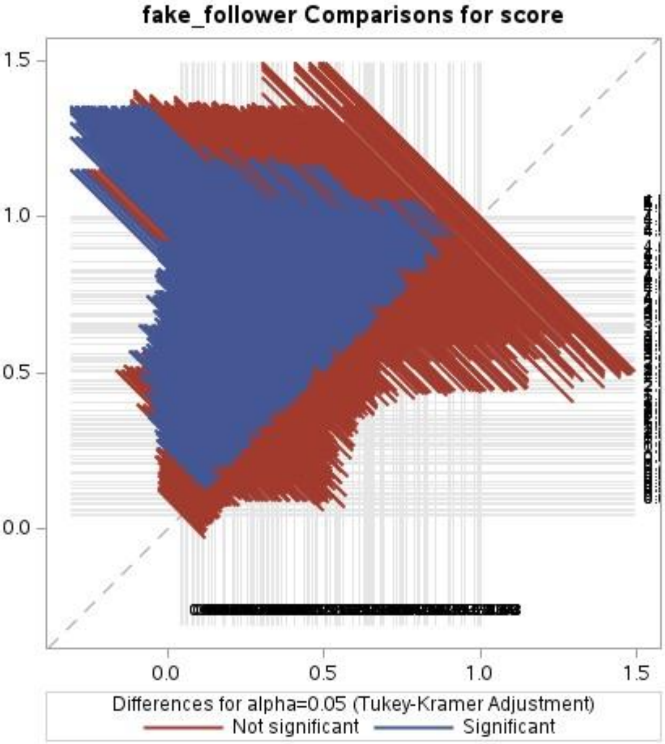


Figure 30 - Turkey-Kramer Adjustment for fake_follower (Covid-19)

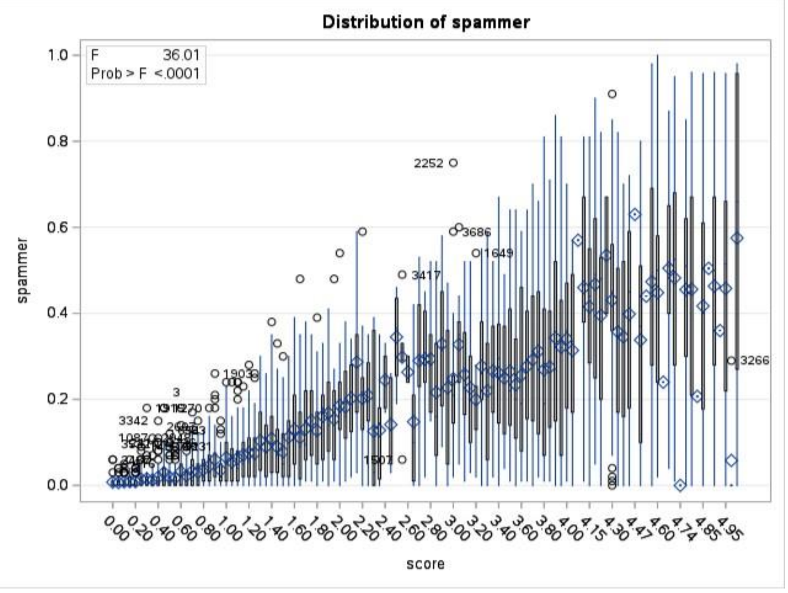


Figure 31 - Distribution of spammer (Covid-19)

Bot Score	Data Min.	Min. Whisker	1st Quartile	Median	3rd Quartile	Max. Whisker	Data Max.	Mean	Standard Deviation	Number of Observations
1	0.00	0.00	0.00	0.01	0.03	0.07	0.26	0.02	0.04	1076
2	0.00	0.00	0.03	0.09	0.17	0.38	0.54	0.11	0.09	748
3	0.00	0.00	0.12	0.22	0.33	0.59	0.75	0.23	0.15	225
4	0.00	0.00	0.13	0.26	0.41	0.83	0.86	0.28	0.19	565
5	0.00	0.00	0.22	0.47	0.63	1.00	1.00	0.45	0.28	1079

Table 11 - Box plot basis - spammer (Covid-19)

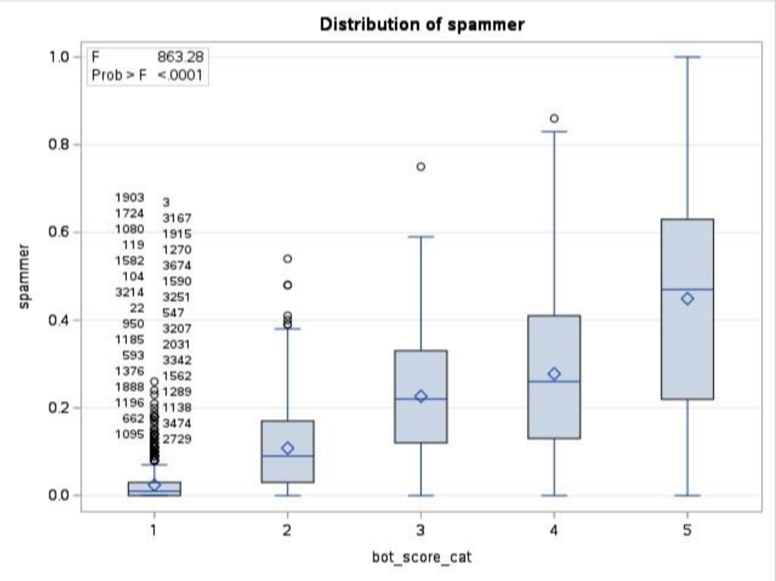


Figure 32 - Box plot – spammer (Covid-19)

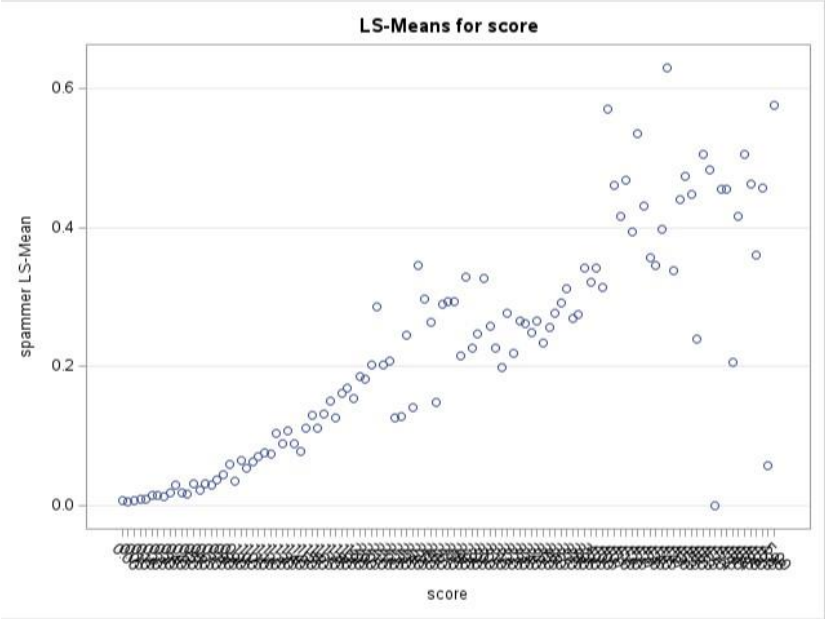


Figure 33 - Spammer Least Square Means for score (Covid-19)

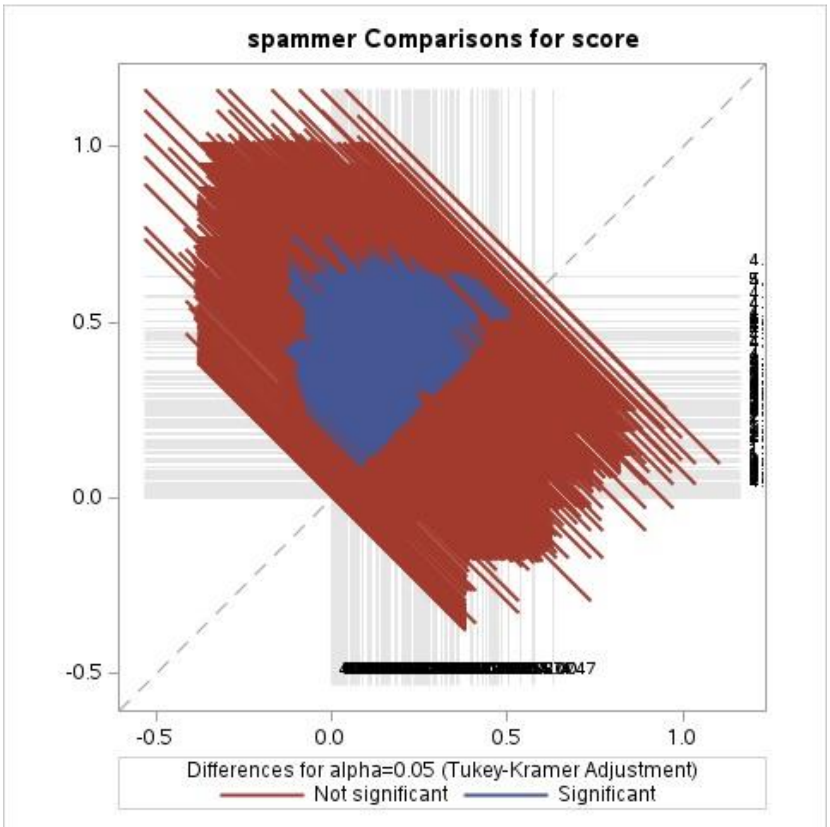


Figure 34 - Turkey-Kramer Adjustment for spammer (Covid-19)

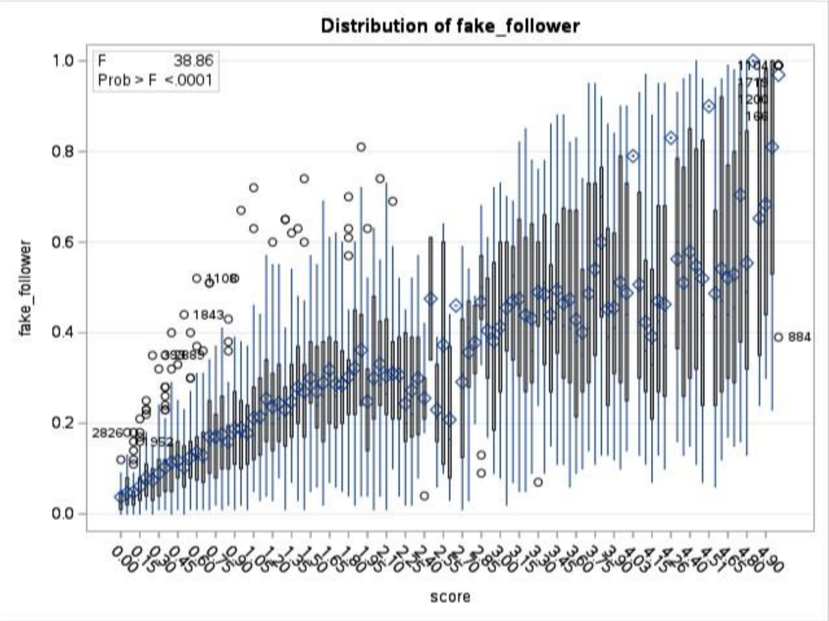


Figure 35 - Distribution of fake_follower (elections)

Bot Score	Data Min.	Min. Whisker	1st Quartile	Median	3rd Quartile	Max. Whisker	Data Max.	Mean	Standard Deviation	Number of Observations
1	0.00	0.00	0.05	0.10	0.17	0.35	0.67	0.12	0.09	1174
2	0.01	0.01	0.17	0.26	0.35	0.62	0.81	0.27	0.14	933
3	0.01	0.01	0.20	0.33	0.46	0.74	0.74	0.34	0.17	245
4	0.02	0.02	0.30	0.46	0.67	0.95	0.95	0.48	0.22	568
5	0.06	0.06	0.30	0.51	0.84	1.00	1.00	0.56	0.29	699

Table 12 - Box plot basis – fake_follower (elections)

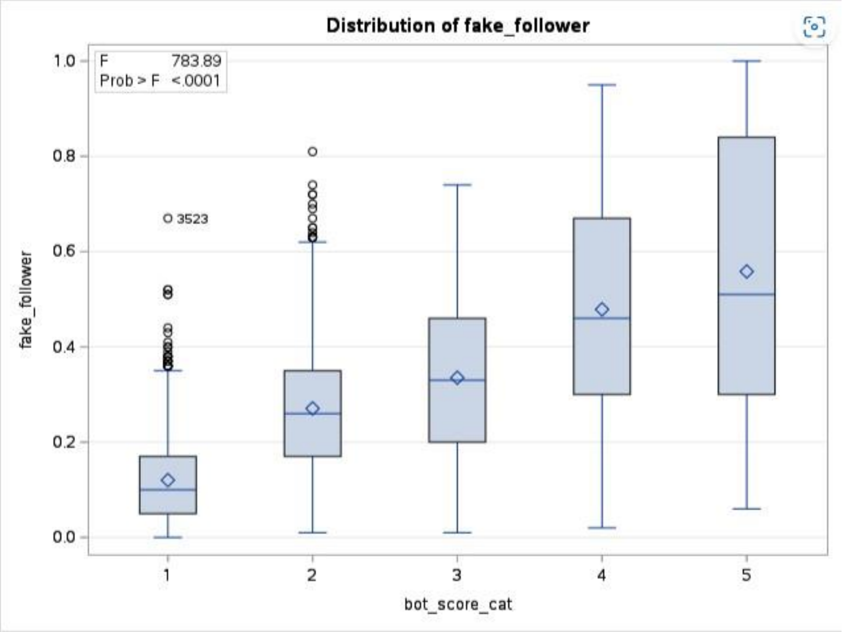


Figure 36 - Box plot – fake_follower (elections)

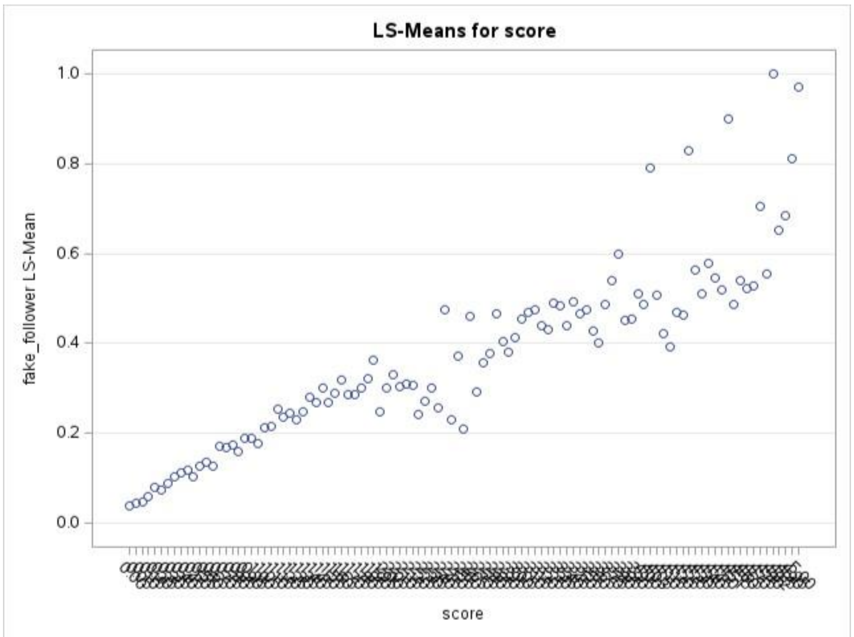


Figure 37 - Fake_follower Least Square Means for score (elections)

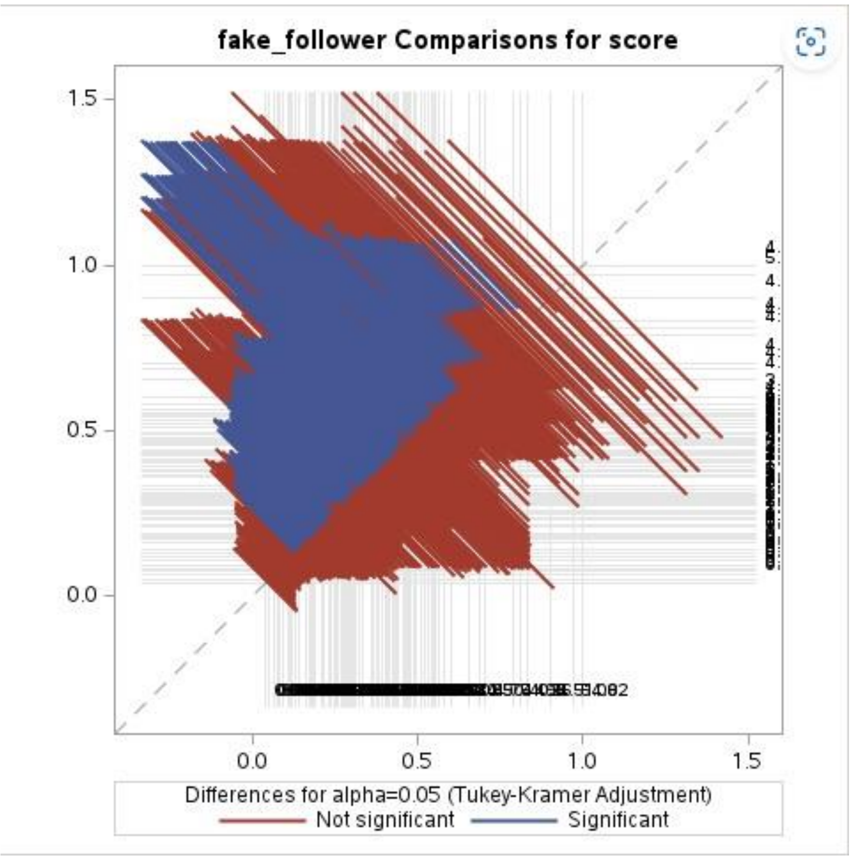


Figure 38 - Turkey-Kramer Adjustment for fake_follower (elections)

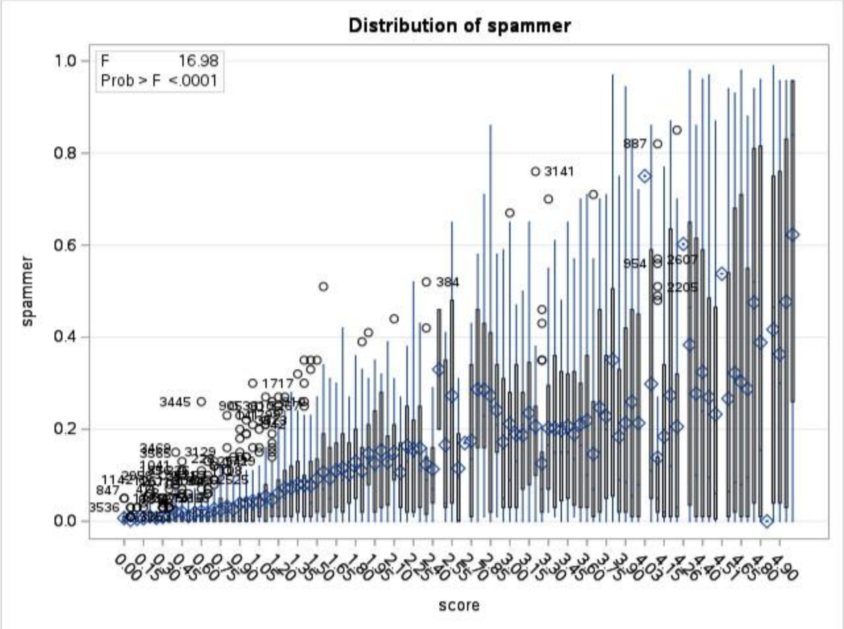


Figure 39 - Distribution of spammer (elections)

Bot	**Data**	**Min.**	**1st**	**Median**	**3rd**	**Max.**	**Data**	**Mean**	**Standard**	**Number of**
Score	**Min.**	**Whisker**	**Quartile**		**Quartile**	**Whisker**	**Max.**		**Deviation**	**Observations**
1	0.00	0.00	0.00	0.01	0.02	0.05	0.30	0.02	0.03	1174
2	0.00	0.00	0.01	0.06	0.15	0.36	0.51	0.09	0.09	933
3	0.00	0.00	0.03	0.15	0.25	0.58	0.86	0.18	0.17	245
4	0.00	0.00	0.03	0.14	0.37	0.87	0.97	0.22	0.22	568

5	0.00	0.00	0.01	0.12	0.64	0.99	0.99	0.32	0.34	699

Table 13 - Box plot basis – spammer (elections)

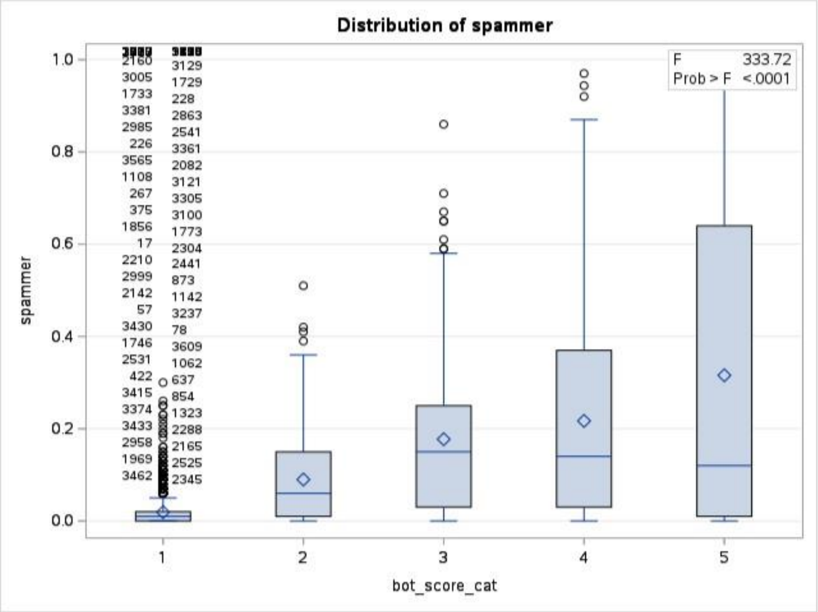


Figure 40 - Box plot – spammer (elections)

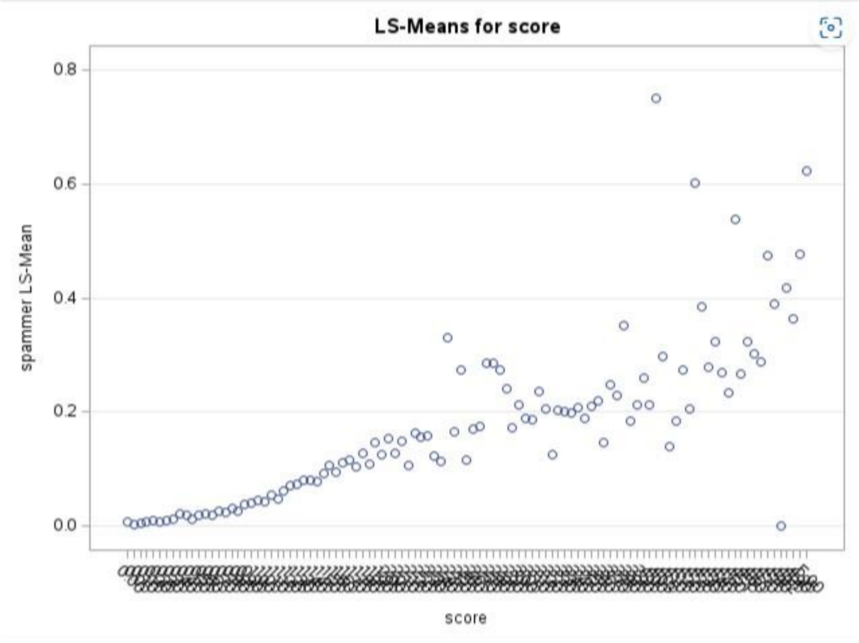


Figure 41 - Spammer Least Square Means for score (elections)

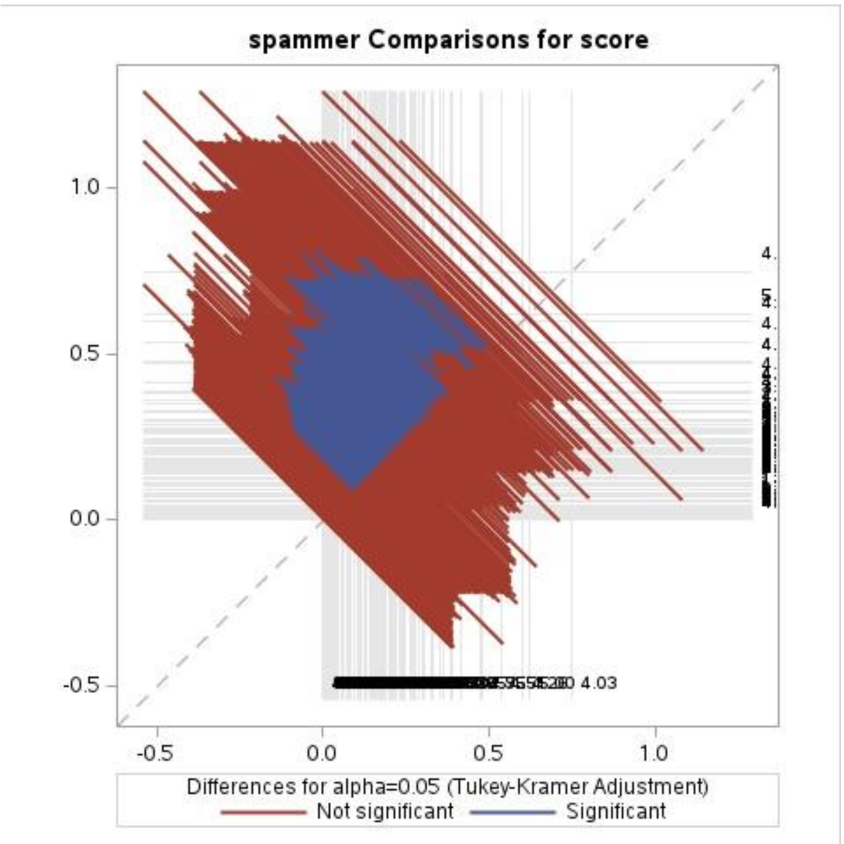


Figure 42 - Turkey-Kramer Adjustment for spammer (elections)

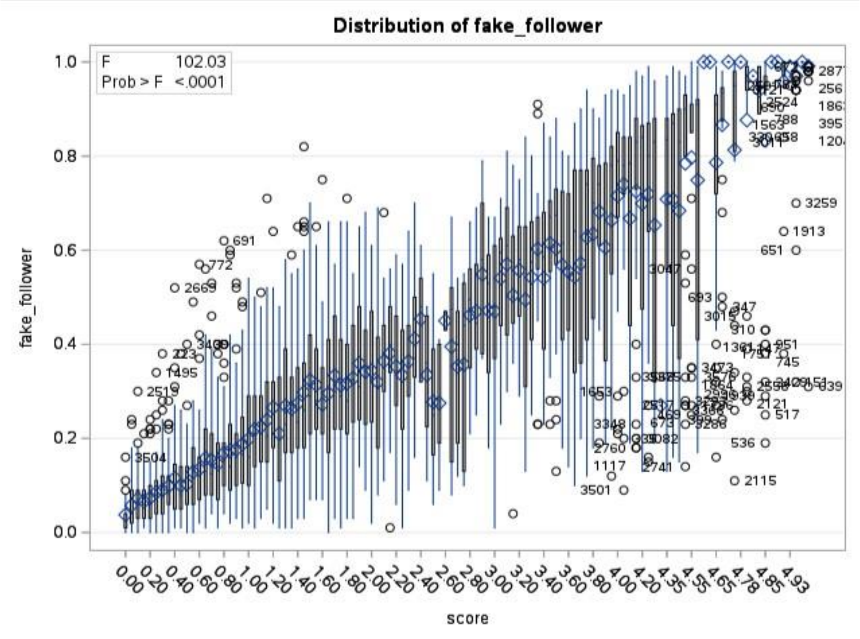


Figure 43 - Distribution of fake_follower (war)

Bot Score	Data Min.	Min. Whisker	1st Quartile	Median	3rd Quartile	Max. Whisker	Data Max.	Mean	Standard Deviation	Number of Observations
1	0.00	0.00	0.05	0.09	0.16	0.32	0.62	0.12	0.10	1171
2	0.00	0.00	0.18	0.26	0.37	0.65	0.82	0.29	0.14	860
3	0.01	0.01	0.30	0.40	0.51	0.79	0.79	0.39	0.16	252

4	0.04	0.04	0.44	0.64	0.75	0.94	0.94	0.59	0.20	561
5	0.09	0.47	0.78	0.91	0.99	1.00	1.00	0.81	0.25	831

Table 14 - Box plot basis – fake_follower (war)

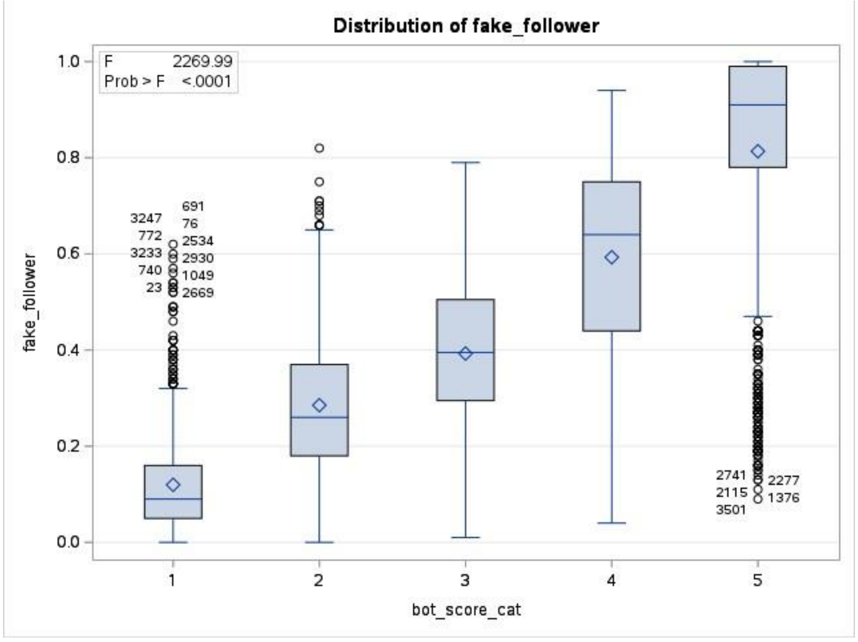


Figure 44 - Box plot – fake_follower (war)

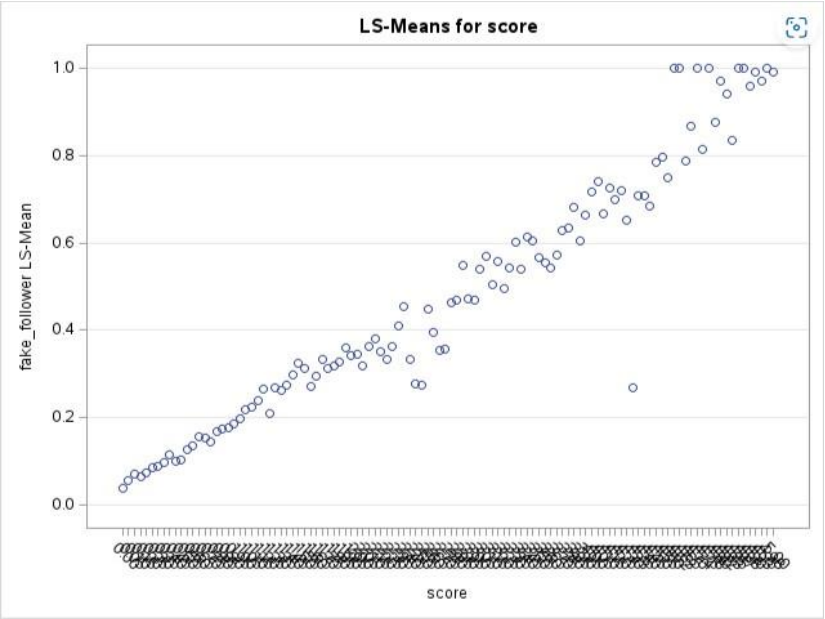


Figure 45 - Fake_follower Least Square Means for score (war)

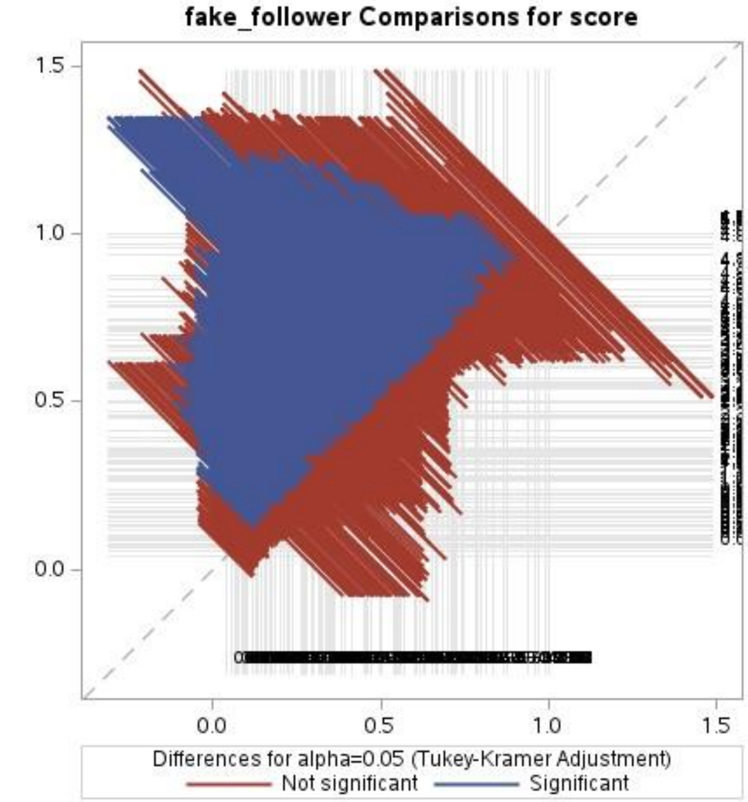


Figure 46 - Turkey-Kramer Adjustment for fake_follower (war)

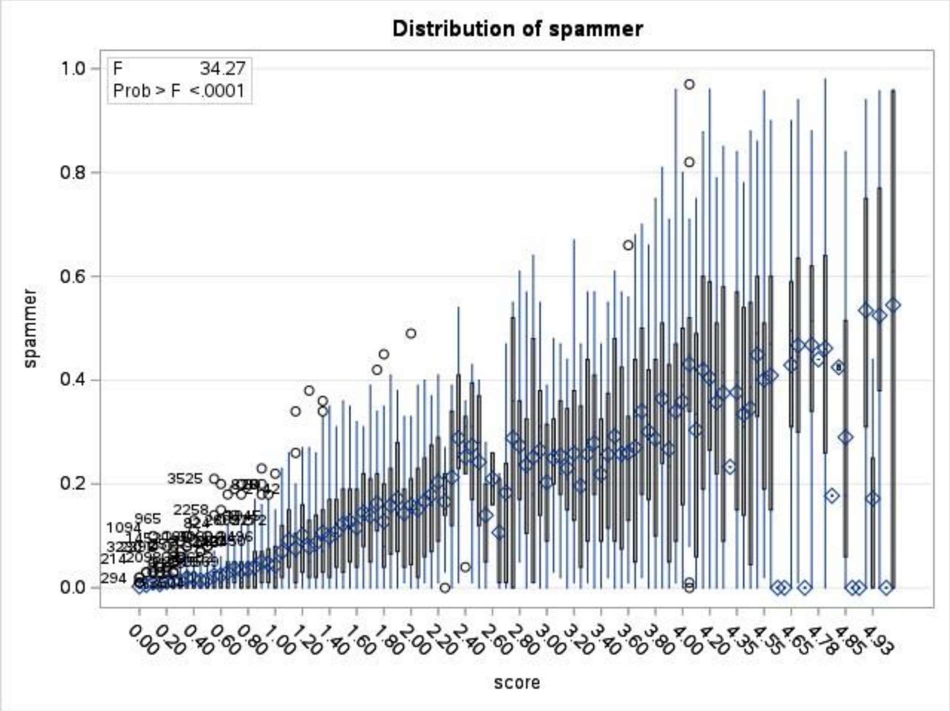


Figure 47 - Distribution of spammer (war)

Bot Score	Data Min.	Min. Whisker	1st Quartile	Median	3rd Quartile	Max. Whisker	Data Max.	Mean	Standard Deviation	Number of Observations
1	0.00	0.00	0.00	0.01	0.03	0.07	0.23	0.02	0.04	1171
2	0.00	0.00	0.03	0.10	0.18	0.39	0.49	0.12	0.10	860

3	0.00	0.00	0.09	0.20	0.32	0.64	0.64	0.21	0.14	252
4	0.00	0.00	0.11	0.28	0.42	0.81	0.96	0.28	0.19	561
5	0.00	0.00	0.18	0.44	0.61	0.98	0.98	0.42	0.28	831

Table 15 - Box plot basis – spammer (war)

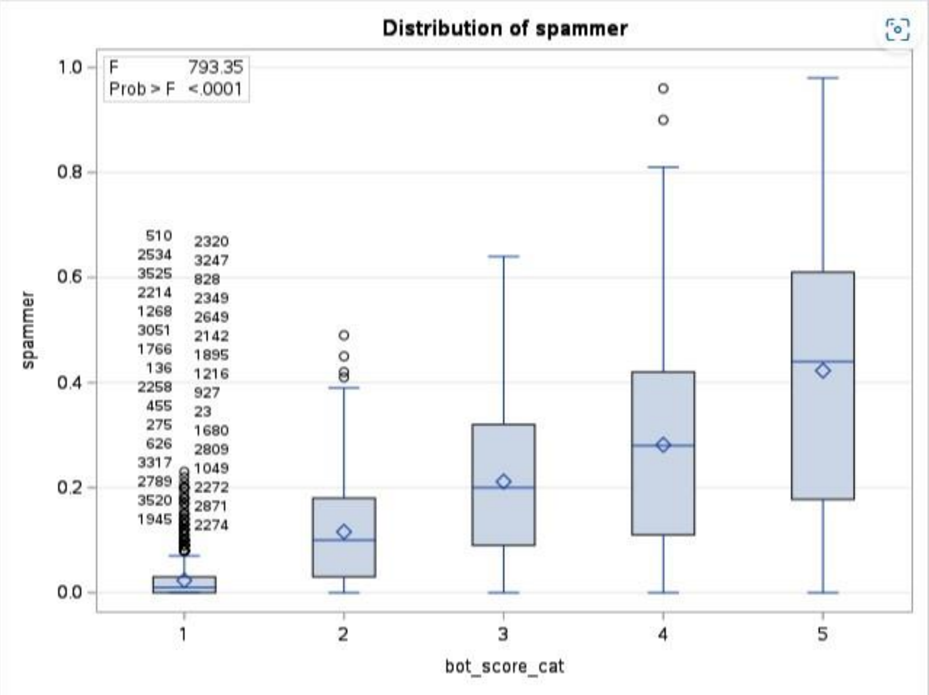


Figure 48 - Box plot – spammer (war)

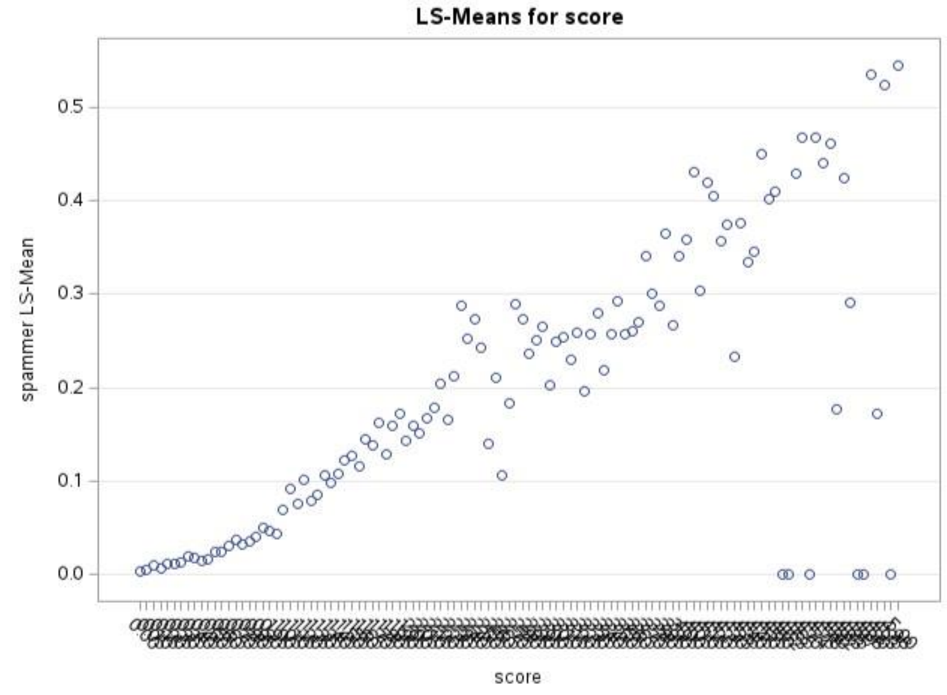


Figure 49 - Spammer Least Square Means for score (war)

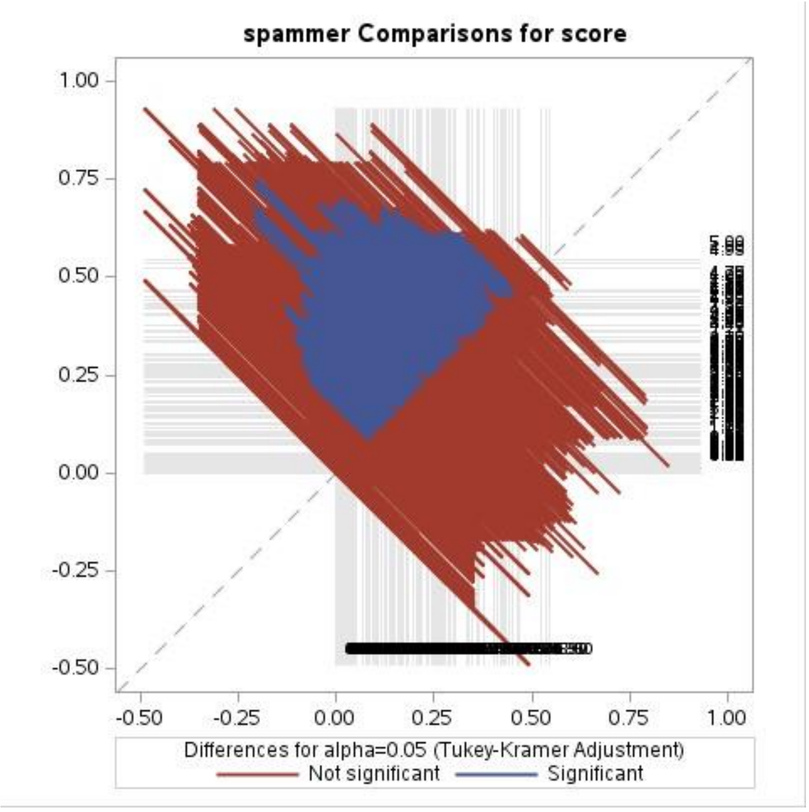


Figure 50 - Turkey-Kramer Adjustment for spammer (war)