Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Information Technologies

Master's Thesis

Detection and Analysis of Twitter Propaganda on Russia-Ukraine War

Ethiopia Gebreselassie

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

DIPLOMA THESIS ASSIGNMENT

Ethiopia Gebreselassie

Informatics

Thesis title

Detection and Analysis of Twitter Propaganda on Russia-Ukraine War

---

Objectives of thesis

The main objective of the thesis is to examine and discover bots on Twitter that were employed during the Russia-Ukraine conflict.

Specific objectives:
- To gather and analyze Twitter data in the context of the Russia-Ukraine conflict.
- To build and verify a statistical model for bots detection in the selected Twitter dataset.
- To evaluate results, interpret findings and formulate conclusions.

Methodology

The methodology of solving the theoretical part of the diploma thesis will be based on the study and analysis of professional information sources. The practical part will include statistical analysis of datasets of tweets containing selected search words about the Russia-Ukraine conflict. The analysis of bots influence will be done by relevant machine learning algorithms. The results will be interpreted and contrasted with other similar studies. 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

80 pages

Keywords

bots, Twitter, search words, Russia, Ukraine, conflict, tweets

---

Recommended information sources

JONES, Marc Owen. The Gulf Information War: Propaganda, fake news, and fake trends: The weaponization of twitter bots in the Gulf Crisis. International Journal of Communication [online]. [Accessed 29 April 2022]. Available from: https://ijoc.org/index.php/ijoc/article/view/8994

KELLNER, Ansgar; WRESSNEGGER, Christian; RIECK, Konrad. What's all that noise: analysis and detection of propaganda on Twitter. In: Proceedings of the 13th European workshop on Systems Security. 2020. p. 25-30.

KHANDAY, Akib Mohi Ud Din, KHAN, Qamar Rayees and RABANI, Syed Tanzeel. Identifying propaganda from online social networks during COVID-19 using Machine Learning Techniques – International Journal of Information Technology. SpringerLink [online]. 29 October 2020. [Accessed 29 April 2022]. Available from: https://link.springer.com/article/10.1007/s41870-020-00550-5

LI, Jinfen, YE, Zhihao and XIAO, Lu. Detection of propaganda using logistic regression. ACL Anthology [online]. [Accessed 18 May 2022]. Available from: https://aclanthology.org/D19-5017/

PETERSON, Austin. Detecting propaganda bots on Twitter using machine learning. Keep Homepage [online]. 1 May 2019. [Accessed 29 April 2022]. Available from: https://keep.lib.asu.edu/items/132431

Expected date of thesis defence

2023/24 SS – PEF

The Diploma Thesis Supervisor

Ing. Miloš Ulman, Ph.D.

Supervising department

Department of Information Technologies

Electronic approval: 14. 7. 2022

doc. Ing. Jiří Vaněk, Ph.D.

Head of department

Electronic approval: 28. 11. 2022

doc. Ing. Tomáš Šubrt, Ph.D.

Dean

Prague on 27. 03. 2024

Declaration

I declare that I have worked on my master's thesis titled "Detection and Analysis of Twitter Propaganda on Russia-Ukraine War" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the master's thesis, I declare that the thesis does not break any copyrights.

In Prague on date of submission _____30/03/2024_____

Acknowledgement

First and foremost, my sincere appreciation goes to my thesis supervisor, Milos Ulman, for his invaluable mentorship, patience, and academic rigor I would also like to thank Czech Government for supporting me with scholarship opportunity to pursue my Master's studies.

I am profoundly thankful to my Lord for the countless blessings, the strength, and the patience bestowed upon me, particularly in times of doubt and stress.

I extend my heartfelt thanks to my mother, Libanos, and my father, Gebretsadik, who have been my pillars of strength and perseverance. Their unwavering love, encouragement, and sacrifice have shaped me into the person I am today. I am eternally grateful for their belief in me and for always being there to light my way forward.

To my boyfriend (Solomon Erkineh), my source of comfort, and my cheerleader. Your confidence in me has been a constant source of motivation and has lifted me during the most challenging periods of my studies.

Detection and Analysis of Twitter Propaganda on Russia-Ukraine War

Abstract

In this digital age, X (Formerly known Twitter) has transformed to become into more that just a social media platform; It has become a key player in the course Russia-Ukraine conflict, where 280 characters can carry the weight of war. This thesis presents a technical exploration into the detection and analysis of X (Formerly known twitter) bot detection on the tweets concerning the Russia-Ukraine conflict. This study used an X (Formerly known twitter) dataset collected in the period of December 2021 to March 2022. This data was obtained from one of the well-known sites, Kaggle. Initial analysis involved descriptive statistics on different columns of the dataset, which provided a general understanding of the data distribution and patterns. The study then progressed into text analysis, extracting textual content to understand the language. A suite of machine algorithms, including Isolation Forest, K-means, and one-class SVM, were employed to identify patterns in the tweets, detect anomalies, and classify instances, assessing each algorithm's effectiveness in detecting anomalies and meeting the research objectives. Ultimately, this study is a map of modern storytelling in a time of conflict. It serves as a guide for future explorers and a call to action for collaborative guardianship of truth in our interconnected world. This work stands as a testament to the enduring power of authenticity and the shared responsibility to maintain it amidst the ever-changing currents of online dialogue.

Keywords: Bots, X (Formerly known twitter), Russia, Ukraine, conflict, tweets, Machine Learning

Detekce a analýza propagandy na Twitteru o rusko-ukrajinské válce

Abstrakt

V tomto digitálním věku se X (dříve známý jako Twitter) proměnil v něco víc než jen platformu sociálních médií; stal se klíčovým hráčem v průběhu rusko-ukrajinského konfliktu, kde 280 znaků může mít váhu války. Tato práce představuje technický průzkum detekce a analýzy detekce botů X (dříve známý twitter) na tweetech týkajících se rusko-ukrajinského konfliktu. V této studii byl použit soubor dat X (Formerly known twitter) shromážděný v období od prosince 2021 do března 2022. Tato data byla získána z jedné ze známých stránek Kaggle. Počáteční analýza zahrnovala popisnou statistiku různých sloupců datového souboru, která poskytla obecnou představu o rozložení a vzorcích dat. Poté studie přešla k textové analýze, extrakci textového obsahu za účelem pochopení jazyka. K identifikaci vzorů ve tweetech, detekci anomálií a klasifikaci případů byla použita sada strojových algoritmů, včetně Isolation Forest, K-means a SVM jedné třídy, přičemž byla posouzena účinnost jednotlivých algoritmů při detekci anomálií a plnění cílů výzkumu. V konečném důsledku je tato studie mapou moderního vyprávění příběhů v době konfliktu. Slouží jako průvodce pro budoucí badatele a výzva k akci pro společné hlídání pravdy v našem propojeném světě. Tato práce je svědectvím o trvalé síle autenticity a společné odpovědnosti za její zachování v neustále se měnících proudech online dialogu.

Klíčová slova: Boty, X (dříve známý twitter), Rusko, Ukrajina, konflikt, tweety, strojové učení

Table of content

1 Introduction 12

2 Objectives and Methodology 13

2.1 Objectives 13

2.2 Methodology 13

3 Literature Review 16

3.1 What is social media 16

3.2 X (Formerly known as Twitter) 17

3.2.1 X (Formerly known Twitter) Bots 18

3.2.1.1 Methods of X (Formerly known Twitter) bot detection 19

3.2.1.2 Deep Learning Techniques 23

3.2.1.3 Graph-based Analysis 24

3.2.2 Forms Of Propaganda 25

3.2.3 **The Effectiveness of Propaganda in Today's New Media Landscape** 27

3.2.4 Propaganda Today 29

3.3 Empirical Studies 30

3.3.1 German Election 30

3.3.2 COVID-19 outbreak and propaganda 33

3.3.3 Gulf Crisis 35

3.3.4 Russian-Ukraine war 37

3.3.4.1 Russian annexation of Crimea in 2014 37

3.3.4.2 Russian invasion in 2022 37

3.4 Summary of the key findings 38

3.5 Research question formulation 39

4 Practical Part 40

4.1 Data collection 41

4.2 Descriptive statistics 42

4.2.1 Tweet Language Distribution 45

4.2.2 Followers Distribution of accounts 47

4.2.3 Verified Accounts Distribution 47

4.3 Text Analysis and Pre-Processing 48

4.3.1 Pre-Processing 49

4.3.2 Analysis of Pre-processed Text 51

4.4 Machine Learning Algorithm Creation for Bot Detection 56

4.4.1 Unsupervised Machine Learning Algorithm Isolation Forrest 57

4.4.2 Unsupervised Machine Learning Algorithm One Class SVM 58

4.4.3 Unsupervised Machine Learning Algorithm K-MEANS 64

5 Results and Discussion 67
5.1 Results 67
5.1.1 Results of Isolation Forrest 67
5.1.2 Results of K-Means 68
5.1.3 Results of One-Class SVM 69
5.2 Discussion 71
5.2.1 Comparison of Pre-War and Wartime Tweets with SVM Algorithm 72
6 Conclusion 74
7 References 77
8 List of tables, figures, code snippets and abbreviations 81
8.1 List of tables 81
8.2 List of Figure 81
8.3 List of code snippets 82
8.4 List of abbreviations 84
9 Appendix 86

1 Introduction

Wardle & Derakhshan, (2017) states that the persuasive impact of social media in moulding the shape of the narratives during conflict times has become a great concern and study over the past few years. As Russia Ukraine war break out, X (formerly known Twitter) has emerged as a key platform where different stakeholders and politicians has engaged in the dissemination of misinformation, amending the power of the public opinion and change narrative discourse.

The conflict, which began in 2014, features a tapestry of geopolitical tensions, territorial conflicts, and a pronounced undercurrent of information warfare. With traditional media wrestling to provide real-time coverage and a plurality of perspectives, According Yaqub et al., (2018), X (formerly known Twitter) has risen as a pivotal forum for narrative control, offering an unprecedented platform for real-time information exchange.

Kollanyi & Howard, (2017)explains amidst all spanning the state-backed disinformation into organic, grassroots in order to actively shape the perceptions, to change the facts and to build ideologically charged viewpoints. X's (formerly known Twitter) fast and vast reach makes it a great platform for such activities, specially during times of conflicts.

In study by Ratkiewicz et al., (2011) mentions that Delving into the tangled web of X (formerly known Twitter) propaganda amid the turmoil of the Russia-Ukraine War is like embarking on a digital detective mission. It's not just about crunching numbers or running algorithms; it's about piecing together a narrative from the pixels and tweets that zip across our screens. We live in a world where a single tweet can ripple across the globe, weaving into the fabric of political discourse. To trace these ripples, to understand how they spread and who's stirring the waters, we need tools—network analysis to sketch out the connections, sentiment analysis to capture the mood, and content analysis to dig into the messages themselves. But it's not just about the what and the how; it's also about the who. In the shadows X (formerly known Twitter) bots—automated agents that can amplify a message or muddle the conversation. Identifying these bots and understanding their impact is crucial. This study sets out to detect and analyse bots and the roles they play in the conflict, utilizing a suite of Machine Learning algorithms to parse the vast troves of data.

2 Objectives and Methodology

2.1 Objectives

This thesis's primary goal is to examine and discover bots on X (Formerly known Twitter) that were employed during the Russia-Ukraine conflict.

Following the main objective of the study, the following are specific objectives raised from the study:

i. To gather and analyze Twitter data in the context of the Russia-Ukraine conflict.
ii. To build and verify a statistical model for bots detection in the selected Twitter dataset.
iii. To evaluate results, interpret findings and formulate conclusions.

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. The practical part will include statistical analysis of datasets of tweets containing selected word searches related to the Russia-Ukraine conflict. The results will be interpreted and contrasted with other similar studies. Based on the synthesis of theoretical knowledge and the results of the practical part, the conclusions of the work will be formulated. Initially the necessary operations to combine the data will be made in python. Along the thesis there will be various packages in python programming language which will be used. Pandas, Numpy, Glob, Langdetect are some of them which will be used very often. To achieve the objectives of the thesis, the intial descriptive statistics will be implemented across the acquired data using different visualizations in python with matplotlib package. It is also important to note that traditional spreadsheet provided by Microsoft which is Excel will also be used to illustrate the important tables. In the descriptive statistics part there will be analysis of categorical and numeric variables depending on the features(columns in other words) of the dataset. The aim will be to get comprehensive understanding of the data to prepare the environment to build the machine learning algorithm at the end. Apart from descriptive statistics, there will be text pre-processing techniques such as tokenization, lemmatization, removal of stopwords and special characters which similarly will be implemented in python. Packages like NLTK and

SPACY offers great tools and existing vocabularies to be able to perform those operations. Awan, A(2023) defines tokenization as a technique of transforming of text into small blocks which are known as tokens in the domain of Natural Language Processing (NLP) and machine learning (ML). The tokens can be long as words or small as a character. This is one of the most important parts which will enable us to work with the words as tokens during machine learning algorithm creations with differnet models. Another important step will be lemmatization which is referring to converting the words to theit initial forms. The usage of same words in different tenses can be eleminated with this way which will be helpful to derive more meaningful results from the text analysis. Last step of the pre-processing will be removing stopwords and special characters as we can assume that propositions and articles have no added to value to the texts generally . Also it is crucial to note that, based on the domain expertise data scientists can generally generate a list of words which they believe do not have any substantial meaning in the context of the study and they can delete them. In our thesis the words which are considered as common sense in the context of Russia-Ukraine war will also be removed. Once the dataset was preprocessed, we conducted descriptive statistics analysis on the most significant columns of the dataset. This analysis provided us with insights into the distribution, central tendencies, and variability of the data. Python libraries such as Matplotlib and Seaborn were used to visualize the descriptive statistics and effectively communicate the key characteristics of the dataset. To gain deeper insights from the textual content of the tweets, we performed deep text analysis. This involved analyzing the language used, sentiments expressed, and topics discussed within the tweets. We leveraged Python libraries like NLTK (Natural Language Toolkit) and SPACY for tasks such as sentiment analysis, named entity recognition, and topic modeling. Visualizations were created using libraries like WordCloud and Plotly to illustrate the findings from the deep text analysis.

In order to achieve our research objectives, we developed and evaluated multiple ML algorithms. Three primary models were utilized: the Isolation Forest, K-means, and One-class SVM. Each model served a different purpose in our analysis.

- Isolation Forest: This model is an unsupervised anomaly detection algorithm that isolates instances in the dataset by randomly selecting a feature and splitting the instances until they are isolated. The anomalies are identified as instances that required few splits to be isolated. We utilized the Scikit-learn

library in Python to implement the Isolation Forest algorithm(Akshahara, 2024).

- K-means: This model is a clustering algorithm that partitions the dataset into k distinct clusters based on their similarity. It operates by iteratively assigning instances to the nearest cluster centroid and updating the centroids until convergence. We implemented the K-means algorithm using Scikit-learn(Sharma, K, 2023).
- One-class SVM: This model is a supervised algorithm used for anomaly detection, particularly when only normal instances are available for training. It constructs a boundary that encompasses the normal instances, detecting any anomalies as instances that fall outside this boundary. We used Scikit-learn to develop and evaluate the One-class SVM algorithm(Scikit-learn, ND).

By employing this comprehensive methodology, utilizing various Python libraries, and developing and evaluating different ML algorithms, we conducted a thorough analysis of the collected dataset. The combination of descriptive statistics analysis, deep text analysis, and ML algorithm evaluation allowed us to gain valuable insights, identify the best-performing algorithm (One-class SVM with features), and make informed conclusions based on our research objectives.

3 Literature Review

The literature study focuses on one aspect of the existing body of literature and highlights the depictions made by eminent authors. As a result, in this literature review is both well-organized and pertinent to the subject of the study. It provides an in-depth discussion of empirical reviews as well as theoretical frameworks, as well as reveals knowledge gaps in the existing body of literature.

3.1 What is social media

According to et al., (2020), social media refers to a group of software-based digital technologies that are usually accessed through apps and websites. These technologies create digital spaces where users can exchange material and information inside online social networks. This inclusive word incorporates prominent platforms such as Facebook, Instagram, and X (formerly known as Twitter), emphasizing the diverse range of features these platforms provide, including content dissemination, messaging, and community interaction. Social media is a highly prevalent and important technology that is utilized by billions of people worldwide. Its importance extends to both personal and professional aspects of life. According to a report by Appel et al. (2020), as of March 31, 2019, Facebook had 2.38 billion monthly active users and 1.56 billion daily active users. It is projected that the worldwide social media user base will reach 3.29 billion by 2022, which would account for 42.3% of the world's population. The extensive audience and their consistent involvement across several platforms highlight the crucial significance of social media in modern communication and marketing methods.

In addition, Dhingra & Mudgal, (2019) delve deeper into the transformative influence of social media on both individuals and organizational advertising strategies. They define social media as a collection of Internet-based apps that utilize the ideological and technological principles of Web 2.0. This facilitates the generation and sharing of material created by users, enabling activities like as collaboration, information exchange, and community participation. Users serve as both consumers and active participants in generating and collaborating on material, encompassing ideas, writings, photographs, videos, and various other forms of media. The influence of this dynamic has made social media an essential ingredient in the marketing strategy of

numerous organizations, extending its importance beyond being an optional part of the promotional mix.

3.2 X (Formerly known as Twitter)

X (formerly known as Twitter) has been recognized by Binsaeed et al., (2020) as the foremost microblogging platform since its establishment in 2006. It has attracted a wide range of users with its features that facilitate blogging, social interaction, and other functionalities. The platform's emphasis on privacy, which allows users to follow people without requiring reciprocal connections, sets it apart from other networks and appeals to a broad audience, including individuals with harmful intents.

The utilization of hashtags on X (Formerly known Twitter) revolutionizes the distribution of content, enabling the mass propagation of both accurate and inaccurate information without the requirement of direct contacts. This signifies a notable transformation in online communication (Binsaeed et al., 2020).

Kumar et al., (2018) found that X (formerly known as Twitter) enables the daily publishing of more than 500 million tweets, which are used for a range of purposes such as sentiment analysis and event detection. The platform's data, encompassing metrics such as tweet shares, favourites, user tweet history, and social connections, play a crucial role in the development of recommender systems. These systems aim to provide users with pertinent and innovative content tailored to their profiles. Bose et al., (2019) acknowledge the substantial volume of user-generated content on X (formerly known Twitter), where users express their opinions and emotions through written comments, thus creating a valuable repository of varied viewpoints.

In their study, (García-López et al., 2011)examine the behaviour of X (formerly known Twitter) users. They find that these users have concise profiles and interact with different types of media. Retweets are identified as the key means by which users disseminate information to a wider audience, extending the reach of news and information on the platform.

3.2.1 X (Formerly known Twitter) Bots

Gilani et al., (2019) explains that a very large number of social networks in online social networks. They are formulated for different purposes, such as distributing news, promoting products, generating links, inflitrating politics, spamming and dissiminating harmful content.

The increasing significance of social networking sites like X (Formerly known Twitter) and Facebook in our daily lives has become evident in recent years, with users heavily relying on these platforms to share information. This reliance has resulted in the widespread growth of social media bots, automated entities that either forward messages or fabricate news, significantly impacting the way information is consumed and shared. X (Formerly known Twitter), with its vast user base posting millions of tweets daily, including text, images, and videos, is notably susceptible to this phenomenon. Luo et al., (2020) estimate that out of X (Formerly known Twitter)'s user base, approximately 48 million accounts are operated by bots. While some of these bots serve relatively benign purposes such as relaying news or system updates, a substantial portion engage in spreading spam or fake news, with potential consequences severe enough to influence political election outcomes.

The significance of social networking sites like X (Formerly known Twitter) and Facebook in our day-to-day lives has grown exponentialy in recent years. Users are susceptible to the messages posted by other users on social networks because they prefer to share their information via social networking sites. This, of course, leads to the proliferation of social media bots, which continually forward messages or fabricate news in an automated fashion. X (Formerly known Twitter) is especially vulnerable to the effects of this phenomenon. In general, users of are able to post millions of tweets each day. Tweets can include not only textual messages but also other rich format messages like images and videos. Luo et al., (2020) estimated that around 48 million of the users who have registered for X (Formerly known Twitter) are bots. Some of these bots merely relay news and automatically update the status of the system. On the other hand, the vast majority of these bots continue to disseminate spam messages or fake news, both of which have the potential to have severe repercussions such as swaying the outcome of a political election.

3.2.1.1 Methods of X (Formerly known Twitter) bot detection

According to Kosmajac & Keselj (2019) Bot (Automated user) is defined as a program that propagated the actual person's behaviour on a particular social network. A bot can work depending upon different parameters like i.e, post likes, user following, tweeting and retweeting. As a whole there two types of bots based on their use: Malicious and Non Malicious. The bots that are honest about their intentions and are not out to cause harm do not have any intention of pretending to be real people on X (Formerly known Twitter). They typically share uplifting phrases or images, tweet news headlines and other useful information, and aid businesses in answering to questions and comments raised by consumers. On the other hand, malicious accounts may generate spam, attempt to get into personal accounnt deatils, trick users into following them or subscribing to scams, to either downplay or amplify political viewpoints, generate trending hashtags for the purpose of financial gain, support political candidates during elections, or generate offensive content in order to troll users. Additionally, certain influencers may deploy bots in order to artificially bloat the size of their audience. This practice is known as "follower inflation." (Kosmajac & Keselj, 2019)

Bot detection can primarily be divided into 4 distinct parts. The first part mainly focuses on Machine learning approches of bot detection with the intention of determining which threats can be located using such techniques and The second subclassifies the 2 main types of machine learning approeches and last two point discuss on the Graph based approach and Deep learning. In the following sub-sections, an attempt will be made to outline these two different streams.

3.2.1.1.1 Machine Learning Approaches

Nguyen et al., (2024) examine the Machine Learning (ML) method, highlighting its significant capability in solving issues related to extensive data sets with multiple variables. Machine learning algorithms can help detect behavioral patterns by analyzing the properties of user accounts. This helps to ascertain whether those accounts are probably managed by bots or humans. (Heidari et al., 2021) further explore the use of machine learning and other advanced techniques for detecting social media bots. They recognize many categories of spam bots, such as promoter bots, URL spam bots, and phony followers. URL spam bots are known for propagating fake URL links by embedding them in retweets of legitimate users.

This approach highlights the intricate nature of bot activity and the need for advanced detection techniques.

The main discussion recognizes that bots disseminate false information, making it difficult to identify purely based on content. Advanced techniques, such as foul language detection or analyzing text properties, can achieve great accuracy in identifying bots. These techniques are crucial because social media bots can focus on different demographics by generating false trends, illustrating the diverse aspects of bot-driven misinformation campaigns, and emphasizing the importance of efficient detection tools.

3.2.1.1.1.1 Supervised Machine Learning Approaches

Initial social bot detectors utilizing Machine Learning algorithms were developed using a supervised learning method. (Nguyen et al., 2024)) states that the detectors have been developed and published since 2010. A standard process for developing a model for a supervised learning detector is outlined in Figure 1

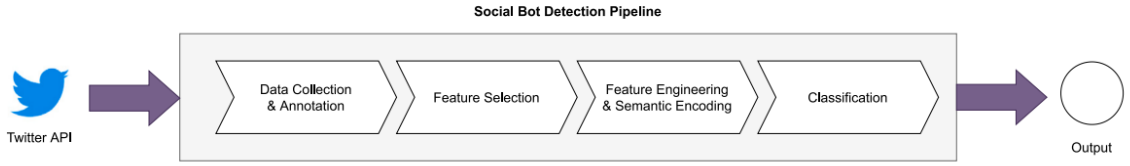

Figure 1 : Pipeline of supervised machine learning detector (Nguyen et al., 2024)

• **Data Collection & Annotation**: A compilation of X (formerly known Twitter) accounts obtained via the X (formerly known Twitter) API, with each item representing a user. Labelling jobs must adhere to a specific schema. Utilizing existing datasets can decrease the time spent in this phase compared to crawling and manually annotating the data.

• **Feature Selection:** Identifying crucial elements for focused analysis from the dataset. Features chosen at this point are mainly influenced by the particular focus or assumptions found in relevant studies.

• **Feature Engineering:** involves converting text and timestamp data into numerical formats for machine learning algorithms, refining the list of features after analysis, and creating new features based on our understanding. Superficial learning outcomes are significantly influenced by this component.

• **Semantic Encoding:** This process involves transforming a user's tweets, denoted as Ts, into a scalar or vector utilizing vectorization or word embedding algorithms.

• **Classification:** selecting the best appropriate architecture to transform the inputs from the previous step into the desired outputs. Feature Engineering and Semantic Encoding are optional in some projects but are commonly used in the development of most detectors.

According to Wu et al., (2022) , supervised learning differentiates between human and bot accounts based on account metrics and behaviour using labelled training data. Researchers have been increasingly using unsupervised methods to address the intricacies of behavioural classification. Supervised learning algorithms necessitate extensive datasets of labelled X (formerly known Twitter) accounts, which might be difficult to locate. This difficulty complicates the development of supervised learning models and may overlook certain bot behaviour.

Aljabri et al., (2023) classify the methods employed in supervised learning for bot detection into three categories: Utilizing methods such as Random Forest, Naïve Bayes, Decision Trees, Logistic Regression, Support Vector Machine, and Neural Networks for Classification, Regression, and Forecasting. These methods utilize diverse data elements like as account usage characteristics, tweet content, and interaction patterns to efficiently distinguish between bots and actual users.

- SVM (Support Vector Machine) is a linear binary classifier that does not rely on probabilities. The algorithm places random points on a graph and then uses a differentiator to separate the classes. The categories are divided according to which side of the line they belong to.
- A neural network employs a Sigmoid function with input, hidden, and output nodes. The input node transmits data which is processed by the hidden nodes, and the output node classifies profiles as fake or real.
- Random forest utilizes numerous decision trees and combines them. It is a blend of many learning algorithms that enhance the overall accuracy.
- The network-centric technique relies on community recognition and egocentric networks, primarily utilized in social interaction characteristics. It focuses on collecting the frequency of tweets rather than the tweet's origin.
- The K_f_reimp technique is a reimplementation of a reference classifier that relies on feature extraction. They have identified the crucial characteristics and utilized the AdaBoost module to develop their classifier.

- The statistical approach utilizes language-independent properties, providing an advantage in accurately interpreting the data. It minimizes the utilization of content parameters.
- The content-based method relies on language-dependent features while disregarding the actual content or subject matter.

3.2.1.1.1.2 Unsupervised Machine Learning Approaches

Ghahramani, (2004) presents the idea of unsupervised learning, in which a machine is provided with inputs (x1, x2, ...)but does not get supervised target outputs or rewards from its environment. This may appear enigmatic, as it is not readily apparent what the machine may learn in the absence of feedback. Unsupervised learning is intended to create representations of input data for tasks like decision-making, predicting future inputs, or effectively transmitting data to another machine. This method involves identifying patterns in the data that surpass what is typically regarded as random noise, with clustering and dimensionality reduction serving as traditional unsupervised learning methods.

Aljabri et al., (2023) classify unsupervised learning techniques into Clustering and Association, which focus on handling unlabelled data. Common techniques in this field are K-nearest Neighbour (KNN), K-means clustering, and Principal Component Analysis (PCA). Semi-supervised learning is introduced, combining a little quantity of labelled data with a big amount of unlabelled data to create a hybrid strategy that benefits from both supervised and unsupervised learning.

Wu et al., (2022) examine the use of unsupervised learning for identifying dangerous spam in Timeline data. They suggest that this method can achieve similar or superior results compared to supervised learning techniques, while reducing human bias, despite the higher level of complexity. The method described in this study identifies malicious spam behaviours such excessive URL shortening, duplicate tweets, and content coordination via unsupervised learning. Chen et al.'s methodology is distinguished from DeBot, another spam detection algorithm, by its superior accuracy in recognizing dangerous spam, especially those that redirect consumers to spam, ad, and malware sites rather than bots linked to news services.

(Rovetta et al., 2020) discuss the benefits of unsupervised learning for analysing online bot activity. It was mentioned the use of session clustering to identify bots. The study utilized

Particle Swarm Optimization (PSO) to differentiate between bots and authentic online users by analysing session characteristics such as total transfer volume, number of pages, and session time. They acknowledge the difficulty presented by the growing proportion of bot traffic, which challenges the belief that bots are just rare occurrences, suggesting that only specific sorts of bots could be identified using this approach.

3.2.1.2 Deep Learning Techniques

Deep learning (DL) is a type of supervised machine learning that uses numerous layers to extract more complex features from the input data. This AI technology replicates the operations and processes of the human brain, as demonstrated by Aljabri et al. (2023). Convolutional Neural Networks (CNNs), Long Short-Term Memory Networks (LSTMs), and Generative Adversarial Networks (GANs) are notable deep learning techniques developed for efficient pattern recognition and data processing.

Hayawi et al., (2023)differentiate deep learning as a distinct subset of machine learning, highlighting its enhanced potential. Deep learning's layered architecture allows it to process and extract features from complex data types such as photos, text, and speech, unlike traditional machine learning methods. Deep learning models have shown better performance compared to traditional, shallow, and machine learning classifiers in tasks like bot identification. Deep learning techniques, namely generative adversarial networks (GANs), have successfully addressed the issue of identifying cyborgs exhibiting human-like behavioral characteristics.

Nguyen et al., (2024) examine how deep learning networks can be used to tackle the difficulties associated with detecting social bots, considering the characteristics of the data. They delineate two main directions in this field: Sequence-Based and Graph-Based Techniques. Sequence-Based Techniques see tweets as sequential data and commonly utilize Recurrent Neural Network (RNN)-based models such as Long Short-Term Memory (LSTM) because of their effectiveness with sequence data. These methods occasionally use Convolutional Neural Network (CNN) models to extract additional hidden features, which improves the accuracy of the results. Graph-Based Techniques describe the connection between tweets and users as a graph and utilize Graph Neural Network (GNN) to extract features. This signifies the most recent method in the field of social bot detection. The discussion recognizes that shallow learning methods are useful, but deep learning has gained

attention for its ability to manage large and intricate datasets without the accuracy limitations seen in older methods.

3.2.1.3 Graph-based Analysis

Nguyen et al., (2024) Social network linkages in graph-based approaches are utilized through methods including trust propagation, graph clustering, and utilizing graph metrics and features. These methods have significant scalability challenges and result in biases. Graph-based methods assume that moving from one social bot account to another is more straightforward, while crowdsourcing results rigorously adhere to the labeling system based on human expertise.

Daya, (2019) Anomaly-based bot detection systems focus on identifying bots by observing and analyzing their behaviors and behavior rather than specifically detecting C2. This approach helps address some of the concerns stated. Graph-based methods reflect host network activities using communication graphs derived from network flows and host-to-host communication patterns. These methods have been suggested in various studies. BotGM constructs host-to-host communication graphs based on network traffic data to represent communication patterns between hosts. Outlier detection is performed using the statistical approach known as the interquartile method. Their findings demonstrate moderate accuracy with few false positives (FPs) across various windowing options. BotGM produces numerous graphs for each individual host. Put simply, a graph is created for each pair of distinct IP addresses. Each node in the graph corresponds to a distinct 2-tuple of source and destination ports, while edges indicate the chronological order of communication. This has a significant overhead and is not suitable for huge datasets. The study employs Machine Learning to cluster the nodes in a graph, emphasizing dimensionality and topological characterization. They assume that benign hosts with similar connection patterns will be clustered together and can therefore be excluded from further research. This significant decrease in nodes greatly reduces detection overhead. Their graph-based characteristics are significantly impacted by severe topological influences. They utilize statistical methods and expert opinions focused on user experience to classify the remaining clusters as either malicious or benign. However, utilizing expert opinion can be difficult, prone to errors, and impractical for extensive datasets. Rule-based host clustering and classification have been recently suggested, utilizing pre-defined thresholds to differentiate between benign and

suspicious hosts. Dependence on fixed thresholds increases vulnerability to evasion and reduces the effectiveness of machine learning-based outlier detection. Graph-based machine learning methods for detecting bots are easy to understand and demonstrate encouraging outcomes. This thesis introduces BotChase, a graph-based bot detection system that can identify anomalies across many protocols. BotChase utilizes graph-based features in a two-phase machine learning method that is resilient to zero-day assaults, spatially consistent, and appropriate for extensive datasets. The process begins with unsupervised learning to decrease the number of training data points in huge datasets, then transitions to supervised learning to achieve precise bot detection.

3.2.2 Forms Of Propaganda

The traditional propaganda is a vertical propaganda. Someone in a position of leadership, personality, or authority who is at the pinnacle of their own prestige will exercise their power to accomplish this. A crowd that has been put in a position of superiority is the audience that will be targeted. The message is delivered from above, and the receiver is expected to remain unengaged. . The direction in which information is spread can be used as a distinguishing characteristic between vertical and horizontal forms of propaganda. In point of fact, throughout history, vertical propaganda has been used as a sort of integration-related propaganda. It is engaged in by the authorities, people, groups of political power, and political power itself. In today's world, the distinguishing feature of propaganda is its use of globalization to disseminate its message through many types of media communications. Horizontal propaganda is distinguished by its tendency to lessen the disparity between the propagandist and the target group. One could argue that it is founded on the principle that all members of the organization are equal. Within a worldwide system of influence, support, promotion, and exposure to the acts of others, each person engages in the practice of propagandizing others and is also the target of others' propagandizing. (Vl\uaduțescu & others, 2014)

The word "propaganda" is most assuredly classified as a member of the "boo" words rather than the "hurrah" words category. It was actually a reference to any term with which we do not agree. It is difficult to have an objective discussion regarding its meaning because it is included in the language of rhetorical insult. The term now carries with it the vernacular baggage of being associated forever in the public mind with the strident polemics of

totalitarian regimes, with World War Two, with Hitler, with Stalin during the Cold War, and in the 20th century both the government of North Korea and Al Qaeda, the worldwide Islamist militant group. This classification of propaganda in extremis serves to confine its operational definition, and as a result, we become desensitized to its finer and more nuanced manifestations. Propaganda is both simple and instructive. (Baines & O'Shaughnessy, 2014a)

Table 1 : Types of propaganda(Baines & O'Shaughnessy, 2014a)

Type of Propaganda	***Explanation***
Propaganda of Enlightenment	Negation of false information
Propaganda of Despair	The inducement of fear of death and disaster
Propaganda of Hope	Presenting to the enemy the hope of a better life if they cease hostilities or surrender
Particularist Propaganda	Seeking to divide the enemy into individual groups and attack them separately
Revolutionary Propaganda	Aiming to break down an enemy from within
Integration Propaganda	Aims at unifying and reinforcing society.
Agitation Propaganda	Aims at fomenting revolution within society.
Atrocity Propaganda	Material containing graphic images of an adversary's savage or barbaric behaviour towards the target audience to arouse their sympathies towards the propagandist.
Sociological Propaganda	The penetration of an ideology into a target audience through its sociological context.
Political Propaganda	The penetration of an ideology into a target audience through its political context.
Vertical Propaganda	That propaganda which makes use of the mass media.

Horizontal Propaganda	That propaganda made by a central organisation which disseminates it for use by small groups.

Plato's greatest concern with rhetoric was that it may make an inferior reason seem like the superior one. This is exactly what propaganda does. The topic of how and in what ways propaganda differs from simple advocacy or a cultural artifact that just so happens to be fashioned around some social or other message is one that is, in fact, still up for discussion. It's possible that one aspect of differentiation is the concept of "intensity" or "commitment." The Sacra Congregatio de Propaganda Fidei was established during the Counter-Reformation by a church that was fighting (to put it crudely) to maintain its market share. This is where the term "propaganda" was first used. As a result, the word carried diverse implications in various nations, with more good connotations in Catholic countries and more negative connotations in other countries due to the nature of proselytizing. The zeal with which a proposition or concept is proposed is suggestive as well as persuasive – 'you must believe' rather than the 'here's why you should believe' typical of marketing, although in the case of the latter, some have argued, particularly in its early years, that modern marketing is itself a form of propaganda.

3.2.3 **The Effectiveness of Propaganda in Today's New Media Landscape**

The challenge of writing about propaganda in history is the same challenge that arises whenever one attempts to write about any communications phenomenon. How can we demonstrate that it is effective, and where can we find objective empirical evidence? The relevance can be easily disregarded due to the fact that there is a lack of persuasive facts. There have been a great number of influential propaganda campaigns, which can all be cited. The well-known "Lord Kitchener" poster, which was used during World War One to help recruit a volunteer army of three million men in the United Kingdom, has perhaps become the most well-known poster in the history of advertising. This specific example does not, of course, prove that the genre as a whole is effective; rather, it only demonstrates the efficacy of a single campaign. However, the impacts of propaganda might extend beyond the clear demographic that it was meant for. (Baines & O'Shaughnessy, 2014a) Communism and fascism were missionary creeds in the 20th century. Proselytizing was central to their practice and meaning. Politics need influence. Facts rarely stand alone. Novel interpretations are possible. To "correctly" understand historical events, historians must examine a political

figure's capacity to deploy myth, symbolism, and rhetoric as a leader skill, not a leadership talent. Winston Churchill's evacuation of the British Expeditionary Forces from Dunkirk is an outstanding example. Using words, tales, and symbols to turn a horrible loss into a victory is one of history's greatest feats.

Baines & O'Shaughnessy (2014) states that Propaganda may be effective because most people are politically indifferent and seek heuristics to make sense of complex political reality. Whether enthusiastically or not, individuals accept public orthodoxy. During the Iraq war, a large part of the American public believed Saddam Hussein was responsible for the 9/11 attacks until investigations proved otherwise. In Great Britain, the government's rationale for war, the public's patriotism, and the troops' success meant that the public supported the troops, even if they didn't favor the war.

In the study (Newsletter, 2018) academics agree, almost unanimously, that propaganda has the potential to be effective. In the field of political science, there is a substantial body of literature that documents the efficacy of various forms of propaganda in altering the views and actions of citizens in a wide variety of settings. However, the mechanisms by which propaganda motivates behavior and the precise conditions under which various types of propaganda are effective remain obscure. Also uncertain are the conditions under which different kinds of propaganda are effective. In particular, while the empirical work on propaganda that consists of factually inaccurate claims and lies has been well-complemented by a considerable body of game-theoretic literature, the theoretical work on other types of propaganda is still an area that is constantly emerging.

Internet technology was praised for its potential to promote democratization and positive political change. This was due to the fact that users were able to use the Internet and "smartphones" to bring attention to international injustices and rally support from around the world in response. Arsenault (2020) debates the decentralization of communication, on the other hand, has made it possible for an increasingly large pool of state and non-state actors to become involved in the dissemination of political 'facts' or 'truths.' This has resulted in an increase in the number of sources that claim to represent credible interpretations of political reality. Internet users have easy access to a wide variety of sources, which enables them to "prove" or reinforce their strong partisan leanings and preferences. This is made possible by

the fact that an ever-increasing number of actors are permitted to participate in the "news" and political debate.(Arsenault, 2020b)

However, the emergence of internet has altered not just the meaning of propaganda but also the opportunities for its use. Everyone has the potential to become a propagandist, and if their message goes viral, they might have an ocean of power at their disposal. It is impossible to overstate the importance of this development; the individual voice is no longer constrained by the difficulty of attracting the attention of the media. A thought-provoking message on YouTube, or even outlandish claims and lies, can just as quickly leave the house of the producer and spread all over the world in an instant. Propaganda spread through the internet is not moderated or screened in any way. Instead of being filtered and digested by a culture's media and review system, lies, fictions, and hatreds are offered in their unprocessed form. The primary area on the internet is defined by the propaganda product, while the secondary space is occupied by contextual criticism in the form of comments that have been "posted," which can be critical or laudatory. As a result, every kind of distortion and false belief has the potential to seep into the larger civic consciousness. This includes conspiracy theory beliefs, like the notion that the CIA or MOSSAD were in some way responsible for 9/11, which refers to the terrorist attacks on the United States carried out on September 11, 2001 by members of Al Qaeda.After the update of web 1.0 to web 2.0 which is driven by users so many journalists have emerged because of this. Student bloggers, for instance, reported live on the massacre that took place at Virginia Tech in April 2007 as the events were happening as they occurred. The uprisings that broke out across the Arab world during the Arab Spring provide a good example of this. Revolutions broke out in Tunisia, Egypt, Libya, Bahrain, Syria, and Yemen (Baines & O'Shaughnessy, 2014)

3.2.4 Propaganda Today

At the beginning of the last century, propaganda was done with leaflets, posters, and print media, which needed large expenditures, resources, and an infrastructure for production and distribution. With radio and TV broadcasting, new types of propaganda emerged and developed, notably due to the informational battle of the XX century's superpowers. Internet and information technology have given propaganda new tools and effect technologies to analyze. (Nazarov et al., 2021)

Propaganda methods are always shaped by their media environment, which includes connectivity, content, and cognitive impact. The Internet and linked gadgets have

revolutionized all of these factors. This new "network propaganda" has the same goals as older kinds (manipulation for political reasons), but exploits the online ecosystem to spread deceptive statements. A growing corpus of work has studied this new ‘computational propaganda' through investigations of ‘trolling,' automated social media accounts (‘bots'), and ‘fake news' or ‘disinformation'.(Till, 2021)

Till (2021) states that The integration of social media information into everyday interactions and mainstream media has expedited the flow between objective, symbolic, and subjective social realities. An individual's view of events ('subjective reality') on social media (whether honestly expressed or not) can provide as content for mainstream media outlets, which then becomes part of the'symbolic reality' that others assume bears some reference to the 'objective reality' (of actually existing social facts). Indeed, social media information can alter 'objective reality' by becoming part of the cultural and political dialogue and requiring answers from political people, giving it political weight (regardless of its veracity)

3.3 Empirical Studies

3.3.1 German Election

In the paper written by Zannettou et al (2019) reports that disinformation campaigns attributed to state-sponsored actors have become increasingly common in the wake of recent political events and elections. In particular, "troll farms," which are allegedly employed by Russian state agencies, have been actively commenting on and posting content on social media in order to further the political agenda of the Kremlin.

The dissemination of propaganda through social media platforms has evolved to become an essential component of cyberwarfare. X (Formerly known Twitter) has been the primary focus of a Russian influence operation that has been directed toward the presidential elections in the United States in 2016. However, the dissemination of propaganda over the internet is not a localized occurrence but rather a problem that affects people all around the world. The effect of political propaganda and fake news is further amplified by journalists who use X (Formerly known Twitter) to acquire "cutting-edge information" when they are chasing down trending topics for their next story and then distribute them via traditional media. This has the effect of making political propaganda and fake news more effective. In this study, we investigate whether a similar influence on political elections can also be

observed in Europe. For this study over 9.5 million tweets thru different hashtags were gathered. The study was mainly on the troll accounts of the Internet Research Agency (IRA). X (Formerly known Twitter) has compiled a list of accounts that are linked to the IRA and had been identified as influential during the process of electing the president of the United States in 2016. This list was done as part of the investigation into Russian interference in the 2016 presidential elections in the United States. In June of 2018, an updated list was delivered to the United States Congress and made available to the general public in order to encourage additional study on the activities of those accounts. In our dataset, we make an effort to identify the same trolls by operating under the presumption that existing X (Formerly known Twitter) accounts are frequently recycled for usage in different contexts.(Kellner et al., 2020)

Table 2 : key words for German election (Kellner et al., 2020)

Party	***Political Direction***	***Term***
Alternative für Deutschland (AfD)	Right-wing to far-right	Afd
Christlich Demokratische Union (CDU)	Christian-democratic, liberal-conservative	Cdu
Christlich-Soziale Union (CSU)	Christian-democratic, Conservative	Csu
Freie Demokratische Partei (FDP)	(Classical) Liberal	Fdp
Bündnis 90/Die Grünen	Green politics	gruene∗
Die Linke	Democratic socialist	linke†
Nationaldemokratische Partei Deutschlands (NPD)	Ultra-nationalists	Npd
Sozialdemokratische Partei Deutschlands (SPD)	Social-democratic	Spd

Following the analysis of a number of tweets included in this research, it was discovered that approximately 79 bots belonging to the Internet Research Agency (IRA) have been attempting to obstruct the electoral process in the United States of America and have also been utilized for the election in Germany.

In the German election, we observe a similar pattern of duplicate retweeting to amplify candidates' messages. Although the influence of protected accounts on RTE size is insignificant in Model 1 (RTEs from the German sample were generally much smaller), it is clear that duplicate retweeters were a powerful driver of RTE scale. In the previous version of Model 1, before we made a decision to control for duplicate accounts, protected accounts were a significant predictor of RTE size (we will take a closer look at those accounts and their behavior below). As seen from Figure 2, all of Alice Weidel's and Sahra Wagenknecht's tweets from our sample were amplified by duplicate tweeters. In fact, this false amplification was the reason why Alice Weidel's RTEs were much larger than those of the other candidates.

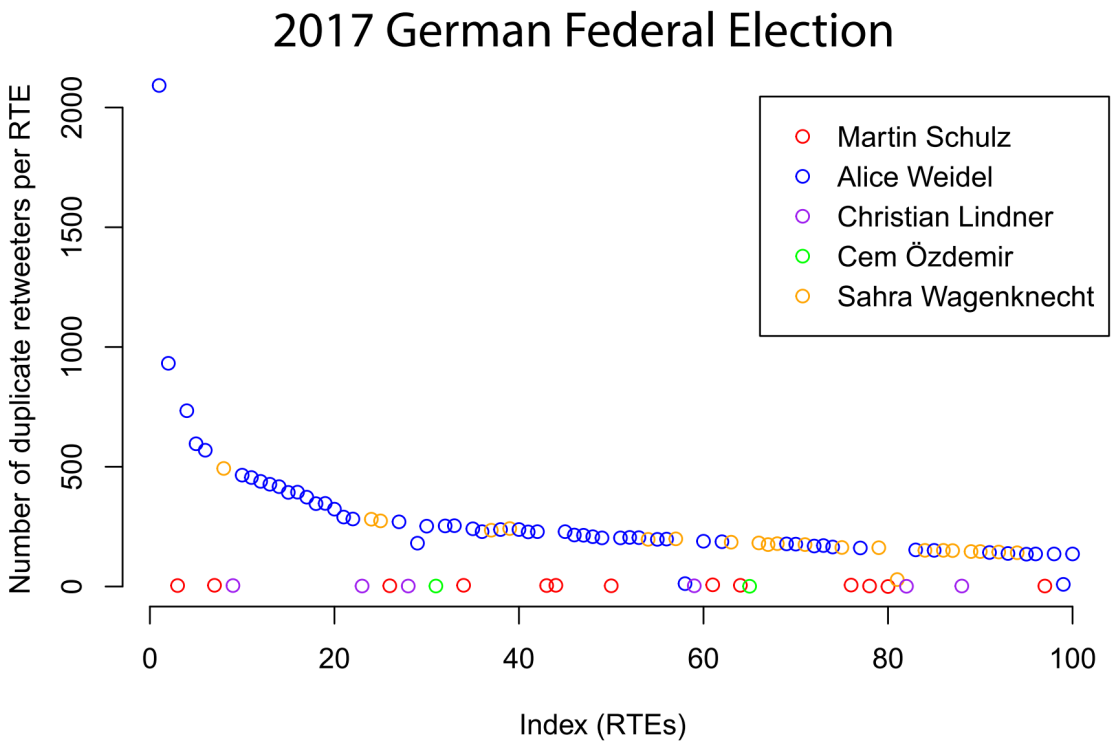

Figure 2 : Amplification patterns in the German election-duplicate retweeters(Boichak et al., 2021b)

Unlike in the U.S. election, in Figure 3 we see that both the protected accounts and the accounts identified as "bots" by Botometer from the German sample have very low followership, but comparatively high tweet rates. Once again, we might speculate some of these accounts might have been involved in duplicate retweeting. These findings suggest that protected accounts were part of a larger orchestrated endeavor, such as botnets that have been previously mentioned in this article. Without speculating whether these accounts were being controlled by a human or an algorithm, we conclude that accounts with a protected

status were important players in the amplification game in both elections.(Boichak et al., 2021a)

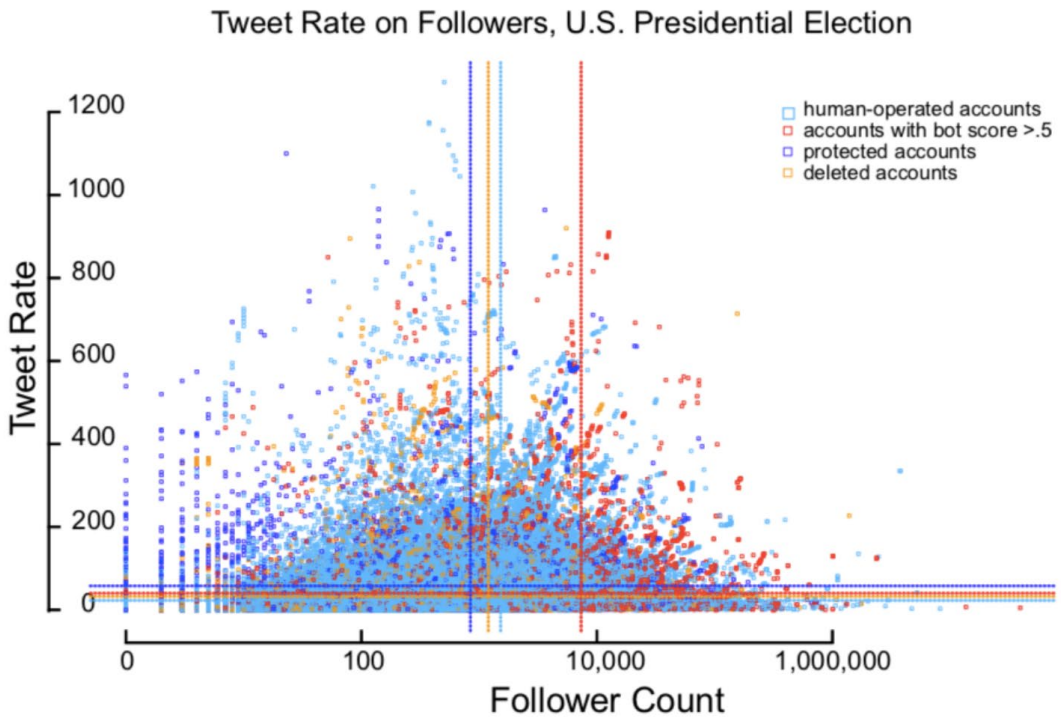

Figure 3 : The distribution of tweet rate by follower count in the US elcetion(Boichak et al., 2021b)

The research conducted by Kollanyi & Howard (2017)used hashtags with different words that are frequently related to the six major candidates. The X (Formerly known Twitter) hashtags analyzed for this study revealed that the AFD-related hashtag accounts for AFD account for 29 percent of the tweets; on the other hand, the tweet hashtags for CDU/CSU account for 18.2 percent; the tweet hashtag in relation to the SPD account for only 8.9 percent; and the tweet hashtag in relation to the SPD account for only 2.6 percent.

Both the research conducted mainly focuses on the German election of 2017, even though the methodologies used by the researchers are different from each other, and the conclusion drawn by both Kollanyi & Howard (2017) and Boichak et al (2021)on the involvement of bots in the election is that even though there were bots created on the course of the election in a slightly increased amount, it did not significantly affect the voters decision.

3.3.2 COVID-19 outbreak and propaganda

Coronaviruses are enveloped, positive, single-stranded RNA viruses that belong to the family Coronaviridae. They are surrounded by a protective membrane. Both the SARS-CoV

and the MERS-CoV are zoonotic in origin; they are responsible for a severe respiratory syndrome; and they frequently result in death. Since the beginning of the epidemic in late December 2019, SARS-CoV-2 has now spread to all continents. As of the 18th of March in 2020, the WHO communicated a total of 179,111 confirmed cases and 7,426 deaths across the globe.(Ciotti et al., 2020)

Because there was no vaccine available for COVID-19, the outbreak spread throughout the entire world. During pandemics, people are more likely to use online social networks due to the increased social distance. Massive amounts of information are being passed about without anyone questioning the reliability of the original source. One of the forms of information that is purposefully disseminated with the intention of obtaining political or religious influence is known as propaganda. It is the methodical and purposeful process of changing an individual's opinion and influencing their thinking in order to achieve the objective that a propagandist has set out to accomplish. During COVID-19, participants are spreading a variety of propagandistic messages about the potentially fatal virus. The data was obtained from X (Formerly known Twitter) by utilizing its application program interface (API), and manual annotation is currently being carried out. For the purpose of selecting the most important characteristics, hybrid feature engineering is carried out. The classification of tweets into one of two binary categories is being done with the assistance of machine learning techniques. Out of all the other algorithms, the decision tree produces the best results. It's possible to get better outcomes by improving the feature engineering, and can be applied by deep learning to help with the classification assignment.(Khanday et al., 2021)

By providing a huge range of tools for transmitting data from one client to another, social networks have helped close the communication gap that previously existed between users. The development of online social networks has made it simpler to exchange information with other people.People use social networking not only to advertizing, marketing and educational purposes but also to acquire ddifferent information about the things that are happening all over the world. Despite all this positive aspects social networking sites can also be a means to spread a misinformation (propaganda). Propagandists are using a variety of current events that are trending in and across the world to their advantage in order to propagate hate, fear, hoaxes, and other forms of misinformation. In late 2019, the city of

Wuhan in China was hit by an outbreak of a virus known as COVID-19. Around the world, over 10 million people were affected by this virus. As a result of countries all over the world engaging in commerce with one another, the virus has spread to every region of the planet, primarily affecting European countries and the United States. This virus has also spread to Iran, India, and Pakistan, among other countries; however, the fatality rate on the Asian subcontinent has remained relatively low so far. A significant amount of investigation is being put towards the creation of a treatment for this pandemic virus. Those who seek to incite fear are exploiting online social networks to disseminate a variety of falsehoods. There is a massive proliferation of false information regarding treatments for this illness. Some of the myths that circulated about how to treat this fatal illness included things like drinking beer and drinking cow pee, both of which have not been demonstrated to be effective treatments by modern medicine. The lawmakers have likewise regarded COVID-19 as a cause for worry. Politicians from different countries all over the world have made an impassioned plea to their constituents to heed the warnings issued by the WHO. Through the use of various online social networks, numerous propagandistic messages are being transmitted. On X (Formerly known Twitter), a number of different hashtags are being utilized in order to disseminate information regarding COVID-19.(Khanday et al., 2021)

3.3.3 Gulf Crisis

Jones (2019) states that Qatar was cut off from the rest of the world by Saudi Arabia, Egypt, Bahrain, and the United Arab Emirates (UAE). This was accompanied by a massive misinformation campaign on social media. This isolation, which some have referred to as "the blockade". The use of thousands of X (Formerly known Twitter) bots was one component of this campaign. The research of X (Formerly known Twitter) bots has typically been limited to the United States of America, Mexico, China, and Russia. This research documents, identifies, and investigates the role that propaganda bots played on X (Formerly known Twitter) during the Gulf crisis. The goal of this research is to fill up these two gaps. The blockade was initially implemented in June of 2017. On May 23, 2017, contentious words were published by the Qatar News Agency, which is run by the state. This was the initial spark that caused tensions to rise. According to reports, these statements were made by the Emir of Qatar, Sheikh Tamim bin Hamad Al Thani. The comments reaffirmed that Qatar has positive relations with a number of foreign nations and organizations, including Iran, the Muslim Brotherhood, and Hamas. In contrast to the customary foreign relations of

the Gulf Cooperation Council (GCC), which hold such organizations and countries in relative contempt—at least publicly—the words stood in stark contrast to those normal foreign relations. During the summit that took place in Riyadh on May 20–21, 2017, King Salman bin Abdulaziz of Saudi Arabia and President Trump of the United States attempted to isolate Iran. Al Thani allegedly made a reference to the significance of Iran as a regional power, which appeared to be a dig at these efforts. When viewed through the lens of the GCC's animosity toward Iran, the comments were understood as Qatar departing from the foreign policy of the GCC, despite the fact that some of these comments would appear to be harmless. Qatar has denied that Al Thani made such comments and has claimed that numerous social media accounts associated with the kingdom as well as the official news station have been hacked. The United Arab Emirates, Egypt, Saudi Arabia, and Bahrain, together referred to as "the Quartet," were quick to dismiss Qatar's claims of a hack as an excuse and organized penalties against Qatar after Qatar's allegations. They accused Qatar of providing support to organizations that were considered to be terrorist groups, such as Hamas and the Muslim Brotherhood then The Qatari publication Okaz led the charge with the title "Qatar Splits the Rank, Sides with the Enemies of the Nation. (Jones, 2019)

There is no doubt that the explicit criticism of Qatar that was broadcast by a variety of state-controlled media channels in the nations that have imposed the blockade was intended to demonstrate that the steps taken by the Quartet were a reaction to a provocation. In point of fact, the four nations were aiming to place themselves on the moral high ground in the situation by putting the focus on the alleged transgressions committed by Qatar. On the other hand, the fury of the media onslaught as well as the intense campaign, along with the longstanding disputes between Qatar and neighboring countries, suggest that the the emir's statements and the purported hack were only a handy pretext on which to hang tensions.

Despite this, the Gulf regimes have not gotten more than a moderate level of attention in the study of propaganda, regardless of whether or not the research has been undertaken online. As new types of harmful conduct continue to surface, it is crucial for the general public to have a greater level of information. This is because it is vital for the general population to have a better degree of knowledge. The use of X (Formerly known Twitter) bots to spread propaganda that was intended to cast a negative light on Qatar and its leadership was the strategy that received the most attention.(Jones, 2019)

3.3.4 Russian-Ukraine war

3.3.4.1 Russian annexation of Crimea in 2014

Russia's annexation of Crimea and attempts to further dismember the Ukrainian state pose a challenge for Russia's neighbors and potentially for the wider European security order of a magnitude greater than anything that has arisen since the end of the Cold War. This is because Russia has attempted to further break up the Ukrainian state by dividing it into smaller and smaller pieces. Given all of the controversy surrounding Kosovo and other conflicts, it is essential to evaluate and refute unjustified legal claims made by Russia in an effort to divert attention from Moscow's use of force and seizure of territory. Doing so is necessary in order to reinforce the principles that underpin European security. In the event that this does not occur, Russia may be ready to stake out a wider legal and normative challenge to western states, going beyond the battles that occurred in the spring and summer of 2014.(Allison, 2014)

After the main Russian TV networks were banned in Ukraine after this year's regime change, Pro-Russian movements turned to social media to spread their cause. Anti-Maidan began as a counter-movement to Pro-European protests in November, funded by Yanukovych and the ruling party. They brought thousands of people from eastern Ukraine to Kiev to show regime support. Anti-Maidan wasn't popular at the time. Protesters were mostly factory workers on paid "vacations" in Kiev. Anti-Maidan's largest group on Russian Facebook equivalent "VKontakte" had 6,000 members in late January. After Yanukovych's collapse, the same page had a quarter-million followers (it now has more than 500,000 and other related pages on different sites including Facebook, gathering even more). This spike corresponded with pro-Russian protests and deadly riots in southeast cities, spurred by online activity. Many opposed closer EU integration..(Unless et al., 2014)

3.3.4.2 Russian invasion in 2022

Chen & Ferrara (2023) analyze the progression of tensions between Ukraine and Russia, highlighting important historical events that led to the conflict. Ukraine declared its independence in 1991 following the disintegration of the Soviet Union. The victory of Viktor Yanukovich, a pro-Russian presidential candidate in 2010, amidst accusations of election fraud, and his removal in 2014, led to increasing tensions. Russia's invasion of Crimea in

2014 exacerbated tensions with Western countries. Russia invaded Ukraine on February 24, 2022, resulting in international criticism, a humanitarian crisis, and a large refugee outflow.

(Fedorenko & Fedorenko, 2022)noted that the Russian war philosophy towards Ukraine in 2022 stems from the lingering sense of loss among the Russian elite following the collapse of the Soviet Union, basically perpetuating the Cold War mentality. Historical reasons for war, known as casus belli, can vary from seemingly legitimate to entirely fabricated, with examples ranging from the Battle of Troy to World War I. They claim that the attempts to create a justification for war on February 22, 2022, in the LPR and DPR regions were not persuasive. This led to Russia's invasion without a formal reason, indicating a change in the contemporary importance of justifications for war.

Garcia & Cunanan-Yabut (2022)examine how social media influences public opinion and policy on the Russo-Ukrainian War in the modern era. As individuals worldwide rely on social media for news and to share viewpoints, public sentiment plays a vital role in influencing policy decisions. International conflict typically remains unaffected by public influence. The war that started in 2014 showed the significance of Russian public opinion in potentially impacting Putin's foreign policy, contrasting with previous research that mainly examined national leaders' utterances. Public opinion in democratic countries has a significant impact and can influence the electoral outcomes of current officeholders. La Gatta et al. (2023) highlight the significance of social media in the war, pointing out that disinformation efforts and the problems of manual fact-checking continue to pose obstacles to preserving informational integrity on the internet.

3.4 Summary of the key findings

Since the emergence of various social media platforms, there has been an upsurge in the distribution of various forms of propaganda relating to political actions taking place in various regions of the world. that at various times and on various circumstances have a more significant impact. The research that were done above illustrate how social media works.

Both the lack of completeness in the majority of the study articles and the fact that the political impact of social media is continuously expanding posed the greatest challenges when attempting to analyze X (Formerly known Twitter) propaganda.

The most typical research gap is that studies are only conducted on X (Formerly known Twitter); as a result, doing broader studies that also include other social media platforms will be an excellent indicator of the effect that propaganda has on a variety of subjects.

3.5 Research question formulation

RQ. How can X (Formerly known Twitter) bots be detected and analysed?

4 Practical Part

In the practical part of our research, we put our methodology into action, conducting a hands-on analysis of the collected dataset. This section aimed to showcase the application of various techniques, methodologies, and machine learning (ML) algorithms to gain insights and make informed conclusions. Through this practical part, we aim to provide a concrete demonstration of our methodology, showcasing the application of data analysis techniques, deep text analysis, and ML algorithms. By following these steps, we obtain valuable insights, make informed conclusions, and contribute to the field of data analysis and ML algorithms in a practical and tangible manner. Below is the Flow-chart which is the brief visualization of what will be done in this section:

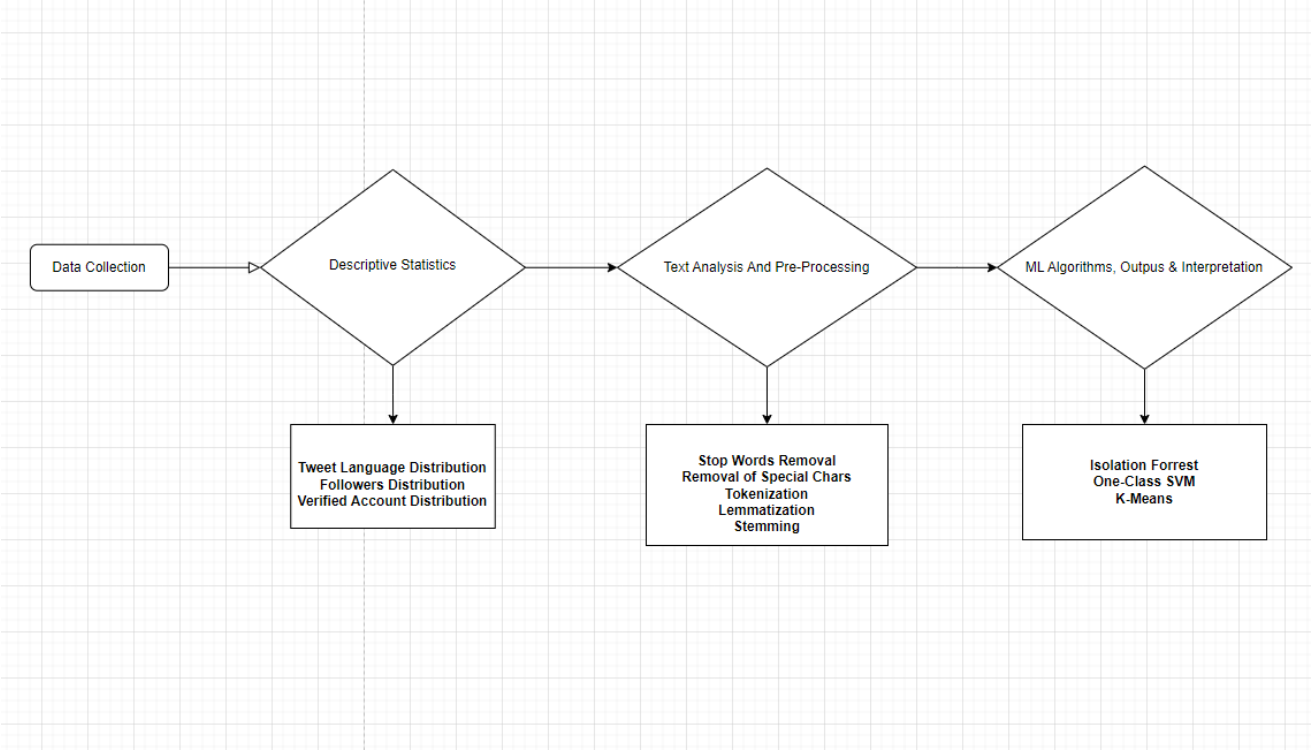

Figure 4 : Schema of Data Analysis workflow

Also it is important to note that in the part where the thesis will demostrate the creation of machine learning algorithms, we will try the algorithms both with features and without features. It is important to list down the selected features which can be potentially helpful to improve the model performance. Below features will all be used in different parts of algorithm wherever it is applicable.

The date of a tweet can be a useful feature in bot detection for several reasons(Date):

- Bots may exhibit regular or periodic posting behavior, such as posting at the same time intervals or during specific time frames. Analyzing the date of the tweet can help capture these patterns and distinguish them from human-like activity.
- Bots may engage in bursty activity, creating a sudden surge in the number of tweets within a short timeframe. By considering the date of the tweet, model can potentially identify these bursts and differentiate them from normal human tweeting behavior.

The number of followers of the account that tweeted a particular tweet can provide valuable information in bot detection (number_followers) :

- Bots can have a significantly higher or lower number of followers compared to a typical human user. Identifying tweets coming from accounts with an unusually high or low number of followers can be indicative of bot activity.
- Bots may employ follower-boosting techniques or engage in follow-back schemes to increase their influence. Analyzing the number of followers can provide insights into the potential use of such strategies.

The verification status of an account, typically denoted by a blue checkmark, can be an informative feature for bot detection(Verification_Status):

- Verified accounts are usually associated with well-known individuals, brands, or organizations. Bots typically do not possess verified status. By considering the verification status, you can identify accounts that are more likely to be genuine and exclude them from bot detection.

4.1 Data collection

To conduct a comprehensive analysis of X (formerly known Twitter) data for The Master's thesis, a meticulously curated dataset was obtained from Kaggle (Russia-Ukraine war - Tweets Dataset (65 days) (kaggle.com)). Kaggle is a well-respected platform that offers a wide range of datasets contributed by the global community of data scientists. It is known

for its reputation in organizing machine learning competitions and serving as a reliable source for high-quality datasets. Specifically, the X (formerly known Twitter) dataset utilized in this study was sourced from Kaggle, enabling access to a substantial number of tweets for in-depth analysis of user behavior, linguistic trends, and other pertinent characteristics.

As part of the data collection process, secondary surveying was employed. The selected timeframe for tweet collection spanned from December 31, 2021, to March 5, 2022. Using Kaggle, a total of 1,316,604 rows of data were extracted, focusing on specific search keywords (not hashtags) including "Russia Invade," "Russian border Ukraine," "Russian troops," "standwithukraine," "Ukraine border," "Ukraine NATO," "Ukraine troops," and "Ukraine war." The search results for each keyword were stored in separate CSV files, resulting in a total of eight CSV files. Each file was named according to the respective search term as follows:

Russia_invade.csv
Russian_border_Ukraine.csv
Russian_troops.csv
StandwithUkraine.csv
Ukraine_border.csv
Ukraine_nato.csv
Ukraine_troops.csv
Ukraine_war.csv

In the subsequent section of the thesis, we will present the descriptive statistics of the entire dataset, providing valuable insights into its overall characteristics.

4.2 Descriptive statistics

In this section, we will explore the descriptive statistics of our dataset before proceeding with any cleaning operations. To achieve this, we employed various Python libraries such as glob, pandas, and matplotlib to merge the separate CSV files, create visualizations, and gain insights from the data. Additionally, MS Excel was utilized for building tables to assist in data analysis.

Initially, we examined the number of columns and assessed the presence of missing values within the dataset. This preliminary analysis provided us with an understanding of the data structure and any potential data gaps or inconsistencies. Using the pandas library, we efficiently explored the column structure and ascertained the extent of missing values.To enhance our comprehension of the dataset, we employed visualization techniques through the use of the matplotlib library. Bar plots, histograms, and scatter plots were created to visually represent various aspects of the data, enabling us to identify patterns, outliers, and relationships.

These visualizations allowed us to gain a deeper understanding of the dataset's characteristics, further contributing to our research objectives. In addition to Python libraries, MS Excel was utilized as a tool for building tables that facilitate effective data analysis. Leveraging the versatility of Excel, we organized and summarized the data, calculated descriptive statistics, and performed additional exploratory analysis. This enabled us to gain valuable insights and extract meaningful information from the dataset, supporting our research objectives. By conducting this analysis of descriptive statistics, we established the groundwork for subsequent cleaning operations. This process ensured that the data is reliable and accurate, providing a solid foundation for our analyses and findings.

Column Name	Number of NAs	Column Names	Number of NAs
_type	0	outlinks	879335
url	0	tcooutlinks	879335
date	0	media	1164571
content	0	retweetedTweet	1316605
id	0	quotedTweet	1179985
user	0	inReplyToTweetId	722546
replyCount	0	inReplyToUser	722546
retweetCount	0	mentionedUsers	643642
likeCount	0	coordinates	1298997
quoteCount	0	place	1298997
conversationId	0	hashtags	979343
source	0	cashtags	1313780
sourceUrl	0		
sourceLabel	0	***Total Rows***	***1316604***

Table 3 : Summary of Missing Data

As we can see from the above table, there are certain columns such as, hashtags, retweeted Tweet and others which contains NA values by vast majority which helps us to sense that those columns potentially may not be useful for our analysis or can be used as a feature in our machine learning algorithm to detect the bots in the upcoming chapters.

It is also crucial to now, the distribution of the search words in order to understand how our data is populated across the chosen words in X (formerly known Twitter).

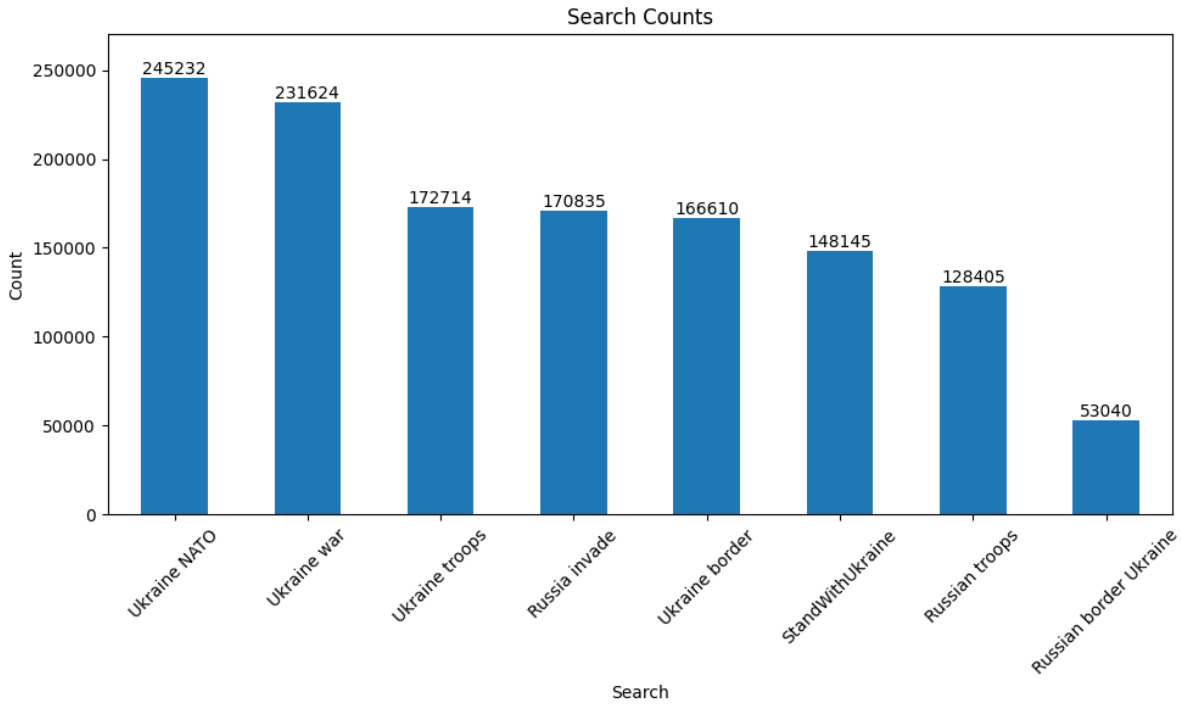

Figure 5 : Distribution of search words

As depicted in the bar chart above, there is a noticeable discrepancy in the distribution of data across the different search terms, with "Ukraine NATO" and "Ukraine war" containing the majority of the data. This imbalance should be taken into consideration when developing machine learning algorithms to ensure fair representation and prevent bias. To address this issue, it is essential to employ stratification techniques during the creation of the machine learning models using scikit-learn libraries. By stratifying the data based on the search terms, we can ensure that the training and testing sets maintain proportional representation from each category. This helps to account for the imbalance in the dataset and can improve the performance and generalizability of the machine learning algorithms. By incorporating the "stratify" parameter in scikit-learn, the models will be

trained and evaluated on a more balanced dataset, mitigating the potential bias caused by the uneven distribution of data. This technique helps to maintain the relative proportions of the search terms, allowing the algorithms to learn and make predictions more effectively across the entire dataset. In summary, the imbalance observed in the dataset, particularly the dominance of certain search terms, should be taken into consideration during the creation of machine learning algorithms. Utilizing the "stratify" parameter in scikit-learn libraries ensures fair representation of all search terms, reducing bias and improving the performance and reliability of the models.

4.2.1 Tweet Language Distribution

The vast majority 92.7% of the tweets in the dataset are in English language, indicating a high level of linguistic consistency among X (formerly known Twitter) users. Secondly, we can see that German is used in approximately 4 % of accross the dataset. The other 3.3 percent of the data is among other languges.

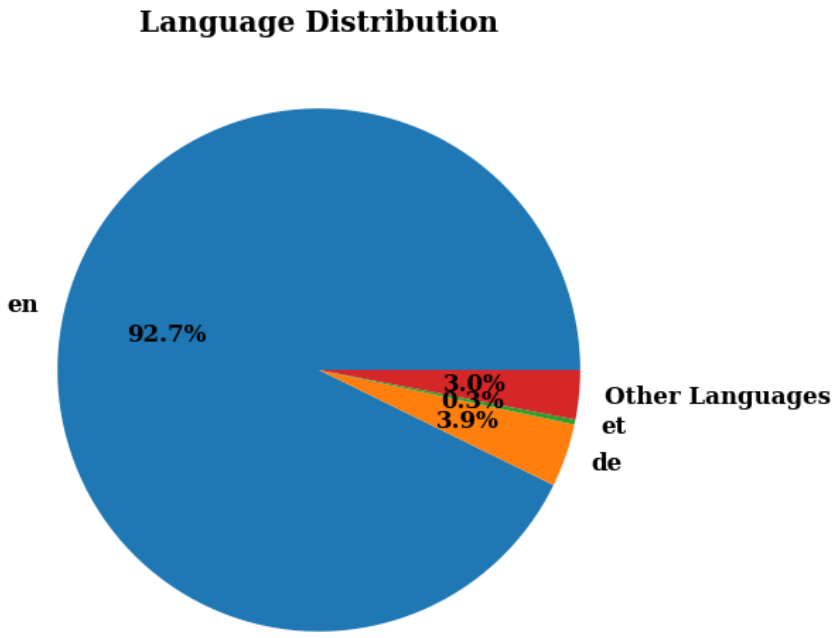

Figure 6 : Language distribution chart

To verify exactly how many languages used in the dataset, the langdetect library in python have been used and a seperate column of language abbreviations was inserted to dataframe.

To be more specific below is the frequency of the lanaguagues appearing accross the dataframe. It is also worthwile to mention that there were 82 observations which langdetect library could not detect the languages and classified it as unknown. Taking into consideration that this is extremely small portion of the dataset we will be able to continue our anlaysis further without those entries. Also the language proportions also help us in making the decision to build our machine learning algorithm to detect the bots in the upcoming chapters as there are many useful methods in python which is working ideally with English language.

Language	Number of Occurences	Language	Number of Occurences
en	1221025	ca	292
de	51531	vi	285
et	4040	tl	272
pl	3988	fa	260
ja	3543	lv	255
pt	3391	el	231
fr	3051	sl	209
it	2794	so	164
af	2349	ta	156
uk	2263	ro	147
es	1970	th	133
tr	1644	sk	123
ru	1549	sq	111
id	1404	ko	83
da	1340	unknown	82
fi	1312	zh-cn	52
nl	1164	mk	51
no	1124	he	41
ar	652	ml	38
sv	643	hu	38
cs	570	cy	33
hr	483	te	32
hi	457	bn	32
bg	399	ur	28
sw	359	kn	21
lt	332	gu	20
mr	9	pa	15
ne	5	zh-tw	10

Table 4 : Language Distribution by Number of Occurrences

4.2.2 Followers Distribution of accounts

In order to achieve this, we had to use the column user which is actually holding a python dictionary inside of it with some valuable data in it. From the dictionary we accessed the element followersCount and used python's lambda function to create a new column which are showing the number of followers belonging to each tweet's account. Once there is the existing column of number of followers, it is possible to build a frequency table for the number of followers:

follower_range	count	percentage
51-100	109197	8.29%
101-200	131227	9.97%
0-50	297041	22.56%
200+	779140	59.18%
Total	1316605	100.00%

Table 5 : Twitter Followers Distribution by Range

Among the X (formerly known Twitter) followers we can see that almost 60 percent of users have more than 200 followers which is actually a good indicator. Taking into consideration that we would expect higher number of followers in the real accounts, we can clearly see that we have many accounts with over 200 followers which might be helpful as a feature later in building our model.

4.2.3 Verified Accounts Distribution

To define the verified and unverified accounts we created another column and named it verification_status. From the user column which is a dictionary of different elements and carries various information about the user, the function in python extracts the value "verified" and check whether it is equal to 'true" or not. If it is true, it assigns the verified_user value to the row as the value and If it is false then we have the value non-verified user assigned. Once the new column is created the pie chart was created to understand the proportion of verified and non-verified users.

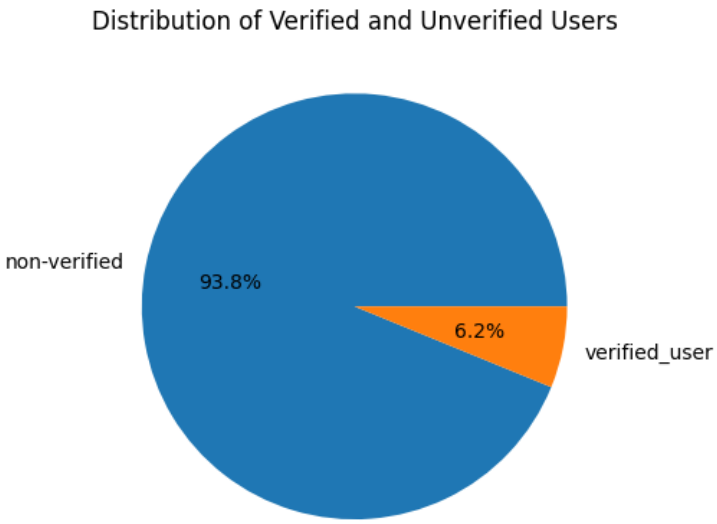

Figure 7 : Distribution of Verified and Unverified users

To be more prices with the numbers it is also important to mention that 93.8 percent of the entire dataset refers to the 1 235 103 number of rows in the dataset whereas 6.2 percent refers to the 81 502 number of rows. It is easy to state the imbalance in the dataset however during the model creation 81 502 rows of data can be really helpful to train our model for the tweets of verified users that we can detect the bots much more accurately.

4.3 Text Analysis and Pre-Processing

In this section we will, we will refer to the content column of our dataset which contains the actual tweets. Along the section we will start with basic text pre-processing which is removing stop words and special characters, tokenization and others. All of those have been covered within the methodology section of the thesis. Later the focus will be on making more visualization of the text data in order to derive valuable insights for extra pre-processing that the data can be ready for a machine learning algorithm to be created. It is important to note that taking into consideration that the majority of the tweets in our dataset belongs to English, the thesis will focus the analysis of the subset of the existing dataset which will contain only the rows in English.

4.3.1 Pre-Processing

Before starting with the pre-processing, it is important to understand the reason behind the it inside of the real scenario. Below is the most frequently used words in the content column without any text pre-processing:

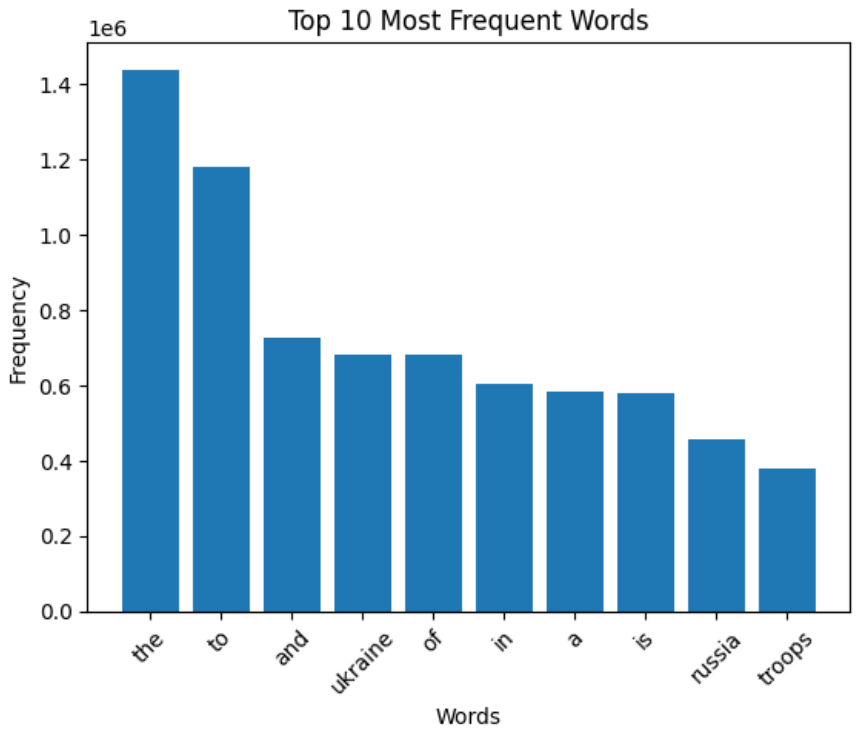

Figure 8 : Top 10 most frequent words

As it can be seen in the above, there are many individual strings which are being considered as words such as propositions. To achieve better results and to make more meaning out of the data we will use the **nltk** library of python to tokenize the words and remove the stopwords from our column. After the mentioned step the most common words

will look like as follows:

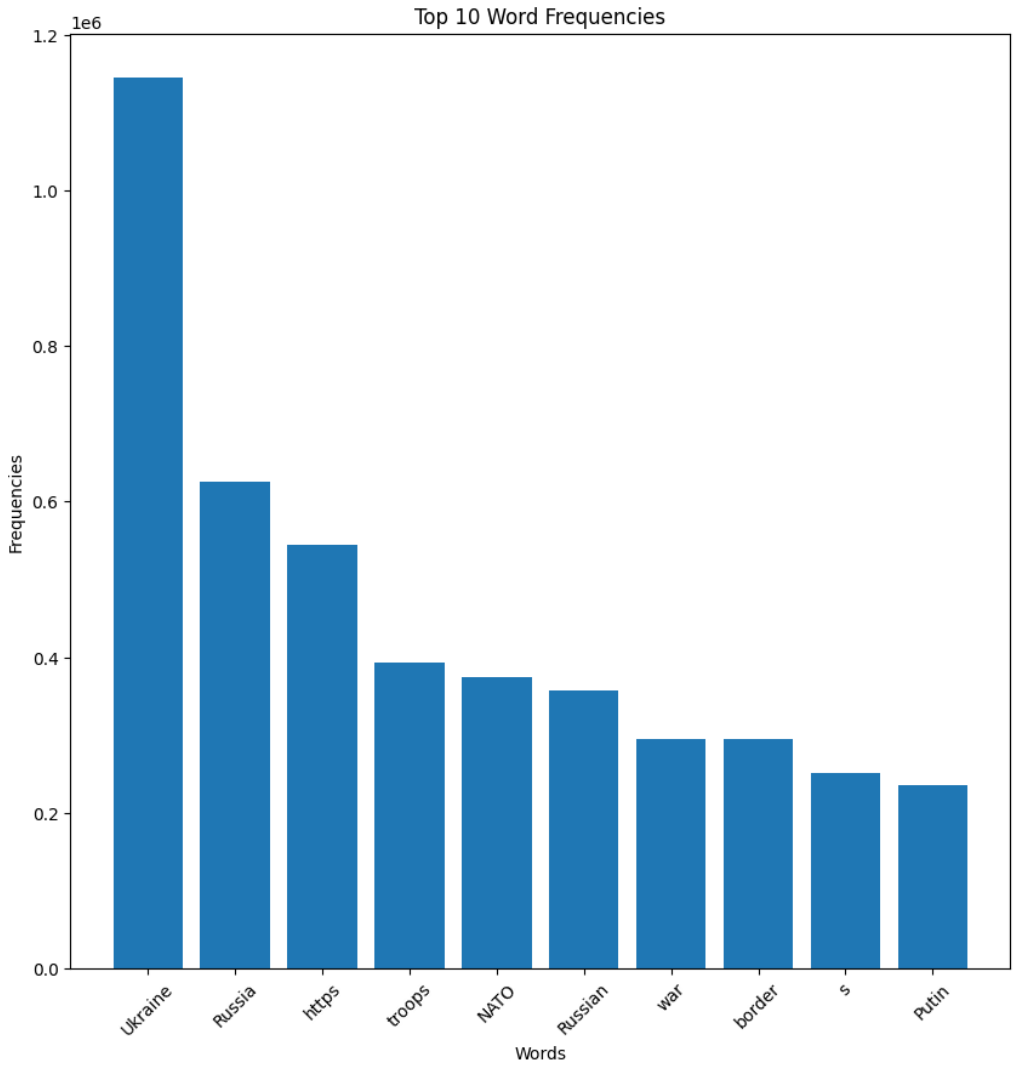

Figure 9 : Top 10 Word Frequencies

After the removal of stopwords, it is seen that the word Ukraine is the most frequent words in our dataset across the content column which follows with Russia. The word https is an interesting point to tackle as it shows how many tweets actually contained certain links within the text.

Taking all above into consideration, in the context of the study it is already obvious that the tweets are about the conflict between Ukraine and Russia. That in mind, it will be useful to create a list of words which are based on our expertise in the field and remove not only stop words such as propositions but also obvious words which do not contribute to the context at any way. Below is the list of words that were decided to be excluded from the

content column across the entire data frame after looking at 50 most frequent used words case insensitive:

'Ukraine', 'Russia', 'Russian', 'war', 's', 'US', 'invade', 'nt', 'would', 'amp', 'StandWithUkraine', 'invasion', 'country', 'Ukrainian', 'like', 'want', 'going', 'Europe', 'think', 'War', 'could', 'world', 'says', 'said', 'one', 'countries', 'get', 'right', 'back', 'know', 'go', 'time', 'President', 'near', 'support', 'need', 'even', 'take'

In the final step of pre-processing, we will refer to the lemmatization and stemming operations where we are bringing all the words into its initial form to decrease the number of similar words and seeing them as one word. The below is the python code snippet which contains the entire pre-processing which have been covered:

```python
import spacy
from spacy.lang.en import English
from nltk.corpus import stopwords
import re
import pandas as pd

# Load the English language model in spaCy
nlp = spacy.load('en_core_web_sm', disable=['parser', 'ner'])

# Download stopwords (optional)
stop_words = set(stopwords.words('english'))

# Custom words to remove
custom_words = ['Ukraine', 'Russia', 'Russian', 'war', 'US', 'invade', 'would', 'amp', 'StandWithUkraine', 'invasion',
                'country', 'Ukrainian', 'like', 'want', 'going', 'Europe', 'think', 'world', 'says', 'said', 'one',
                'countries', 'get', 'right', 'back', 'know', 'go', 'time', 'President', 'near', 'support', 'need', 'even', 'take']

# Function to remove stopwords, special characters, and lemmatization using spaCy
def preprocess_text(text):
    # Tokenize the text
    doc = nlp(text)

    # Remove stopwords, special characters, and custom words
    filtered_tokens = [re.sub(r'[^\w\s]', '', token.lemma_.lower()) for token in doc if token.lemma_.lower() not in stop_words.union(custom_words)]
    filtered_tokens = [token for token in filtered_tokens if len(token) > 1]  # Remove individual characters

    return ' '.join(filtered_tokens)

# Apply the preprocess_text function to the 'content_cleaned' column
df_english['content_preprocessed'] = df_english['content_cleaned'].apply(preprocess_text)

# Print the updated DataFrame
print(df_english)
```

✓ 56m 15.1s

Code Snippet 1

In the next steps, the expectation from the visualization and analysis can raise and more valuable insights from the data can be acquired.

4.3.2 Analysis of Pre-processed Text

One of the useful techniques to gain further insights into the content of the tweets is by examining the most frequently used words. To accomplish this, we can employ the

wordcloud object from Python's wordcloud library and generate a wordcloud visualization for our pre-processed text.

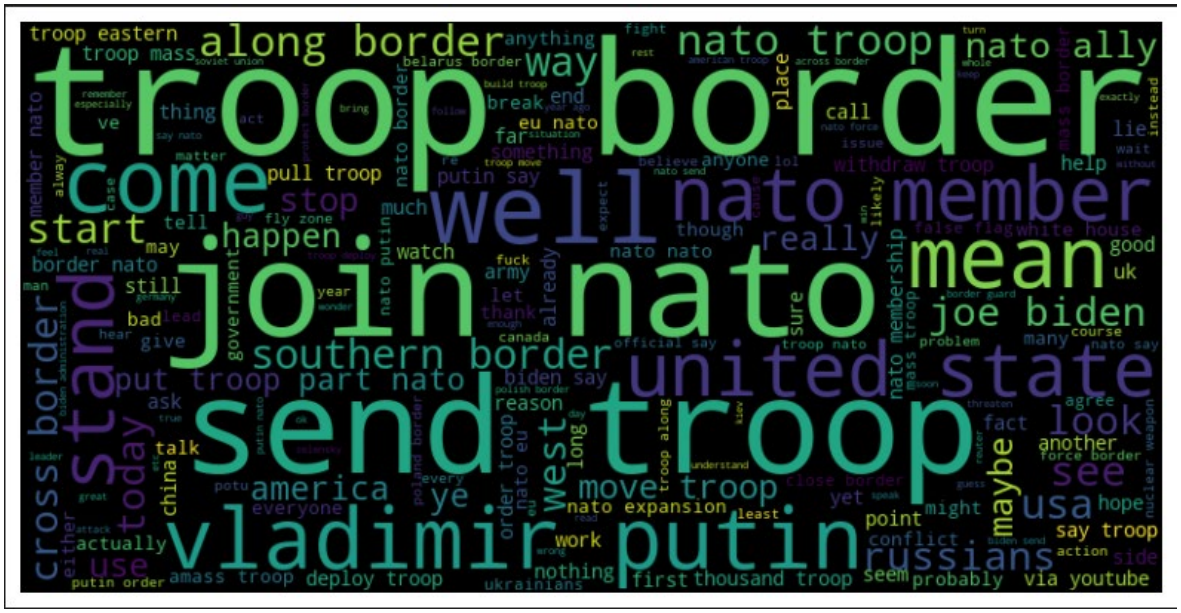

Figure 10 : Word Cloud Library

As it can be seen above, there are certain words which already give some insight about the content of the tweets. It is possible to sense that most of the content is about the troops that moves within the region as we can refer to the words such as “move troop” “troop border”. Especially the geographic locations can be falsely tweeted intentionally to confuse the enemies. Also, there are words like “NATO expansion” which potentially can be the around of propaganda context against NATO by Russian bots or even actual people.

To be more accurate with the information it is also handy to look for the most used words in the pre-processed texts:

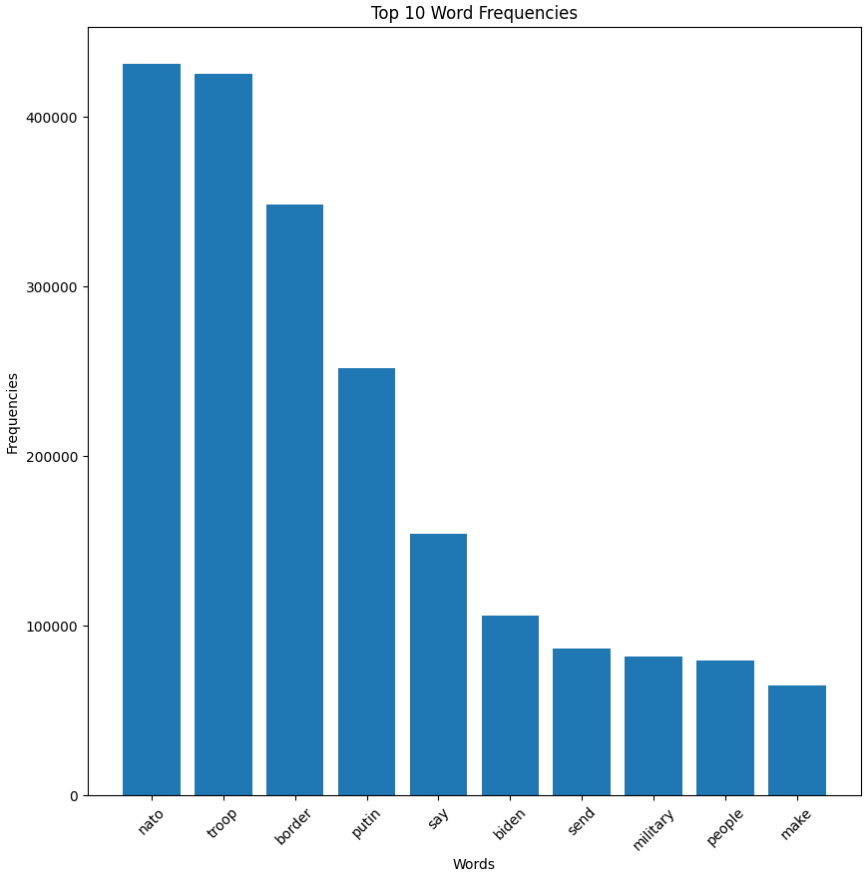

Figure 11 : Top 10 word frequencies in pre-processed texts

One more time, troop is one of the mostly used words which again raises suspicions about the potential use of bots or false information that both political sides use to mislead each

other. Furthermore, it is important to visualize the most used words by verified accounts in order to be able to anticipate the potential differences between bots and actual users:

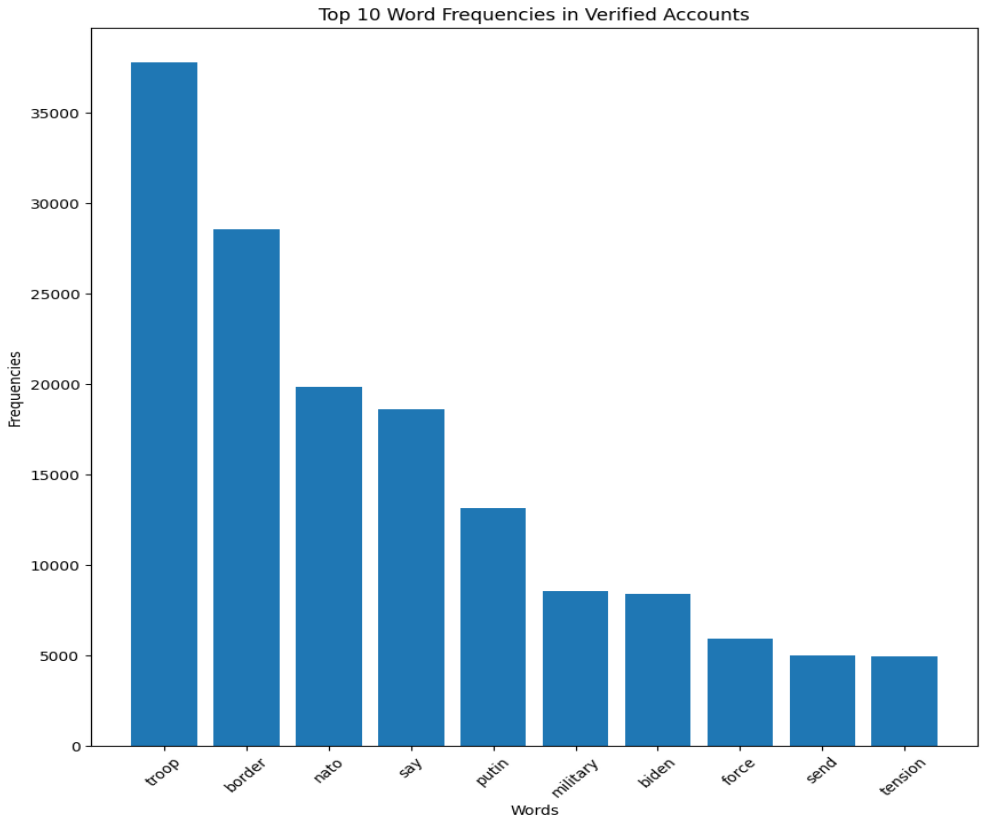

Figure 12 : Top 10 word frequencies in Verified Accounts

As it can be seen above the words are pretty similar in verified users which looks like a balanced texts in both verified and unverified users.

Another interesting point of view is visualizing the relationship between word frequency and sentiment and to interpret it:

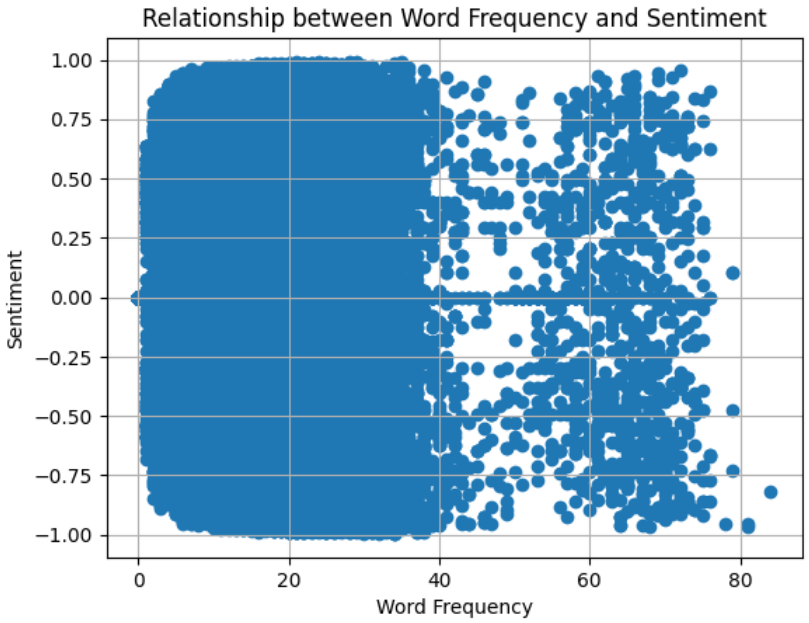

Figure 13 : Relationship between word frequency and sentiment

As per the observations an approximate equal distribution of sentiment scores between -1 and +1, and the frequency values are distributed evenly between 8 and 40, we can derive the following interpretations.

- Balanced Sentiment: The equal distribution of sentiment scores around -1 and +1 suggests a balance between positive and negative sentiment in the text data. It indicates that the sentiments expressed in the content are evenly spread between positive and negative, resulting in a relatively neutral or balanced sentiment overall.
- Moderate Word Frequency: The even distribution of word frequency between 8 and 40 indicates that the words appearing in the text data occur with similar frequencies. This suggests that the words are used consistently throughout the content, without any specific words being significantly more or less frequent than others.
- Consistent Tone: The combination of balanced sentiment and moderate word frequency implies that the text data contains a relatively consistent tone. The even distribution in both sentiment scores and word frequency suggests that the content expresses a range of sentiments without any specific words dominating the discourse.
- Contextual Analysis Required: It's important to note that the interpretation may vary depending on the context and the specific domain of the text data. While an evenly distributed sentiment and word frequency can imply balance and

consistency, a deeper qualitative analysis of the content is necessary to understand the underlying meanings and themes.

4.4 Machine Learning Algorithm Creation for Bot Detection

This section will cover the various machine learning algorithm creations which we will discuss in results and discussion section. Taking into consideration that there was not have any prior training dataset given, there could be 2 approaches to achieve the desired outcome.

First way would be using an unsupervised machine learning algorithm such as anomaly detection with Linear SVC or Clustering Algorithm and then carefully analysing some of the tweets to anticipate the performance of the algorithm based on our domain knowledge. To achieve this, some of the other columns such as account_status can be used as features because it is directly linked to real users. This will be covered in the next sub-chapter.

The second way will be deciding to use verified users as real tweet owners. What can be done is to split the dataset of verified users, train and test the model based on them. Later on, using the confusion matrix the best performing model can be chosen and applied to the unverified users. However, it is important to take into consideration that the date is huge and getting a sample dataset out of it will be the main approach in the context of this thesis.

To achieve the ideal number of sample size we will use the Qualtrics.com as they are providing a free open-source sample size calculator based on the desired confidence level, population size and margin of error. Those indicators will be as below.

Confidence level: 99%
Margin of error: 1%
Population size: 1221024
With the above indicators the ideal sample size is calculated as 16355

Also, to create best performing model, out of 16355 rows, it is important to keep the same proportion of verified/unverified users as it is going to be crucial during the stage of using features and also for readers to compare the difference between bot and non-bot. Below is the python code on how to achieve that sample size:

```
import pandas as pd

# Randomly select 16355 rows from the "verified" users
df_verified = df_english[df_english['verification_status'] == "verified"].sample(n=818)
# Randomly select 16355 rows from the "non-verified" users
df_non_verified = df_english[df_english['verification_status'] == "non-verified"].sample(n=15537)

# Concatenate the two dataframes to form the final dataframe with the desired proportion
df = pd.concat([df_verified, df_non_verified])

# Shuffle the rows of the final dataframe
df = df.sample(frac=1).reset_index(drop=True)
```

✓ 3.2s

Code Snippet 2

4.4.1 Unsupervised Machine Learning Algorithm Isolation Forrest

Initially we can start with one of the most popular algorithms which is Isolation Forest. Here we are setting the anomaly threshold to -0.5. Choosing an appropriate threshold is important as it determines the trade-off between false positive and false negative rates. Setting a more negative threshold would result in classifying more instances as anomalies, potentially including more actual bot activities. Conversely, setting a less negative threshold would classify fewer instances as anomalies, potentially missing some bot activities. That is the reason we are choosing the generally accepted value of -0.5 and try our model with content_preprocessed:

```python
from sklearn.ensemble import IsolationForest
from sklearn.feature_extraction.text import TfidfVectorizer
import pandas as pd

# Convert preprocessed text data into numerical feature vectors
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(df['content_preprocessed'])

# Train the Isolation Forest model
isolation_forest = IsolationForest()
isolation_forest.fit(X)

# Predict the anomaly scores
anomaly_scores = isolation_forest.decision_function(X)

# Set a threshold to classify instances as bot or non-bot
threshold = -0.5  # Adjust the threshold based on your data and requirements

# Create a new column 'bot_label' in the DataFrame with the predicted labels
df['bot_label'] = ['Bot' if score < threshold else 'Non-Bot' for score in anomaly_scores]

# Print the instances and their bot labels
for index, row in df.iterrows():
    print(f"Instance {index + 1}: {row['content_preprocessed']} | Bot Label: {row['bot_label']}")
```

Code Snippet 3

The results are not very promising as it could not detect any text as bot. Taking this into consideration instead of trying to improve the model with features moving to the One Class SVM might be more logical.

4.4.2 Unsupervised Machine Learning Algorithm One Class SVM

One of the most popular NLP algorithms for unsupervised data in anomaly detection can be used which is One Class SVM. Initially considering the pre-processed content column it is important to train the model based on the only that column. The model then will decide on the un-verified users where there is an anomaly or not. Below is the code snippet on how to achieve that outcome. In One-Class Support Vector Machines (One-Class SVM), the parameter p represents the probability estimate for an instance to be considered abnormal or outlier. The p parameter is related to the fraction of training instances that are expected to be outliers or anomalies. It controls the proportion of the data that is estimated to lie outside the model's target region. More specifically, p determines the tolerance of the algorithm for false positives.

```python
from sklearn import svm
from sklearn.feature_extraction.text import TfidfVectorizer
import pandas as pd

# Convert preprocessed text data into numerical feature vectors
tfidf_vectorizer = TfidfVectorizer()
X = tfidf_vectorizer.fit_transform(df['content_preprocessed'])

# Train the One-Class SVM model
ocsvm_model = svm.OneClassSVM()
ocsvm_model.fit(X)

# Make predictions on the same feature vectors
predictions = ocsvm_model.predict(X)

# Map the numerical predictions to labels
labels = ['Normal Instance (Unverified User)' if p == 1 else 'Anomaly Detected' for p in predictions]

# Update the labels based on the verification status
for index, row in df.iterrows():
    if row['verification_status'] == 'verified':
        labels[index] = 'Normal Instance (Verified User)'

# Create a new column 'anomaly_label' in the DataFrame with the predicted anomaly labels
df['anomaly_label'] = labels

# Print the instances and their anomaly labels
for index, row in df.iterrows():
    print(f"Instance {index + 1}: {row['content_preprocessed']} | Anomaly Label: {row['anomaly_label']}")
```

✓ 59.6s

Code Snippet 4

By using the condition p == 1, the code snippet checks if the score p for a given instance is equal to 1. If the score is equal to 1, the instance is classified as a "Normal Instance (Unverified User)". This means that the One-Class SVM model considers the instance to be within the normal range of the training data, possibly representing an unverified user.

As a result using the value_counts() method of python, the following number of labels are assigned by the algorithm:

Labels	Numbers
Normal Instance(Unverified User)	7634
Anomaly detected(Unverified User)	7903
Normal Instance(Verified User)	818

Table 6 : Number of Anomaly detected and Normal instance detected with One class SVM with out Feature

Now, it is crucial to have a look on some of the anomaly detected texts and normal instances in order to understand the difference on decision and anticipate whether the algorithm does the common sense or not wihtout any features selected. Below are some examples from anomaly detected instances and normal instances which will be discussed (*All classified items can be also found in the attached appendixes which contain the result*

of the algorithms in pandas dataframe in columns in csv format).

Anomaly detected by algorithm:

- *troops will 'fight until the very last breath' lol I HEAR the same WARNIG from the Afghanistan GOV you remember? and 3 days later they give the KEYS of the country to TALIBAN looks like Ukrainians taking the same ROAD as the camel riders*

- *Volunteers cross Polish border into Ukraine to fight Russian forces\n\nKonstantin Shukhnov and Corky Siemaszko\n\nThu, March 3, 2022, 5:04 PM\n\nAwesome!*

- *@BrouwerRudolf @Freeminded1987 @iPicNews What do you mean aura, Ukraine Poland and Baltic coutries are free now and No one can decide about their inner politics. NATO is about Defence not offence.*

- *@LeftistAfrican @bibk_o @rammy_c8 Russia: Moving troops inside their country which is totaly legal = wArMoNgErInG\nNato: sending troops to the rusian border and actively threaten russia = ThEy ArE jUsT dEfEnDiNg UkRaInE*

Looking at the first tweet, it is possible to have a suspicioun on the anomaly which have been labeled by the model. The tweet contains some hate speech about Taliban and Ukraine which potentially can be part of a propaganda.

From the above results, it is also possible to analyze some of the anomalies which are incorrectly detected. The second tweet contains the crossing, checking on the names Konstantin Shukhnov and Corky Siemazsko, Corky is a reporter in and successfully found the same news in NBC (NBC News, [2022]) which is in the following link: [https://www.nbcnews.com/news/world/volunteers-cross-polish-border-ukraine-fight-russian-forces-rcna18619.]

Similarly, 3rd and 4th tweets look more like a personal opinions rather than propaganda as it does not contain false informative message. Furthermore, it is easy to note that it is a reply to multiple people which makes it more real to think that it is a bot.

Normal Instance by algorithm:

- *Photos from the huge #StandWithUkraine rally in NYC Times Square today. Stay strong, @Ukraine!\n\nList of rallies, all over the world, and other links: https://t.co/Mo4cuW15gR\nTheir Twitter account: @RazomForUkraine https://t.co/sQiwwtD4bR*
- *Mr Putin talking up nuclear around Ukraine, NATO and an eventual battle for Crimea, sucking in everyone. World leaders have not learned anything in last 100 years about pandemics or wars. Very little winners and lots of losers in all scenarios. Always created by politicians.*
- *@BoHines Ukraine is not a member of NATO. This is a political war 🇺🇸opportunity for puppet Biden. His Vietnam 🇻🇳 https://t.co/rHyiCdacsD*
- *@Palestine616 Russian Troops are moving to Donetsk*

In the above normal instances there are also some of the tweets which indeed can be consiedered as normal instance from the first glance however the tweet which shows the directon of troops moving to Donetsk can be intentional misinformation to mislead the enemy and affect their planning.

In the next step, the features will be used to implement the same algorithm and look to the results. This time we will use, date, verification_status and also the followers_count as features due to their importance in being able to assume the propaganda bots. Bots generally would follow a time pattern and expectation would be to have less followers in those accounts. That is the reason we expect better results adding those columns as features. Below will be the updated code in order to achieve that. In the mean time as we know that the verified users have real tweets, we will already give them the label of Normal Instance:

```python
# Convert preprocessed text data into numerical feature vectors
tfidf_vectorizer = TfidfVectorizer()
X_text = tfidf_vectorizer.fit_transform(df['content_preprocessed'])

# Encode the 'verification_status' column as numerical labels
label_encoder = LabelEncoder()
df['verification_status_encoded'] = label_encoder.fit_transform(df['verification_status'])

# Convert the 'date' column to numeric representation
df['date_numeric'] = pd.to_datetime(df['date']).astype('int64')

# Select the 'followers_count', 'verification_status_encoded', and 'date_numeric' columns as additional features
X_additional = df[['followers_count', 'verification_status_encoded', 'date_numeric']].values

# Concatenate the TF-IDF matrix and the additional features
X_combined = hstack([X_text, X_additional])

# Train the One-Class SVM model
ocsvm_model = svm.OneClassSVM()
ocsvm_model.fit(X_combined)

# Make predictions on the same feature vectors
predictions = ocsvm_model.predict(X_combined)

# Map the numerical predictions to labels
labels = ['Normal Instance (Unverified User)' if p == 1 else 'Anomaly Detected' for p in predictions]

# Update the labels based on the verification status
for index, row in df.iterrows():
    if row['verification_status'] == 'verified':
        labels[index] = 'Normal Instance (Verified User)'

# Create a new column 'anomaly_label' in the DataFrame with the predicted anomaly labels
df['anomaly_label'] = labels

# Print the instances and their anomaly labels
for index, row in df.iterrows():
    print(f"Instance {index + 1}: {row['content_preprocessed']} | Anomaly Label: {row['anomaly_label']}")
```

✓ 1m 22.0s

Code Snippet 5

As a result, the following table can be created based on the value_counts() method of python(*the full results can be found in attached appendixes of pandas dataframe as ipynb*):

Labels	Numbers
Normal Instance(Unverified User)	7929
Anomaly detected(Unverified User)	7608
Normal Instance(Verified User)	818

Table 7 : Number of Anomaly detected and Normal instance detected with One class SVM with Features added

As previously done, it is significant to look into some examples to understand what algoritm have chose as anomaly. Some of the normal instances and anomalies are as follows.

Anomaly detected:

- *#Ukraine:in the centre of satellite image,a small clearing is visible and on the road heading towards #Kyiv a convoy of Russian military vehicles 40 miles long and 15,000 troops. It's an attempt to put Kyiv under siege.https://t.co/mGQ6JD78Pc #RussiaUkraineWar #russianinvasion*
- *US troops on alert amid threat of Russian invasion in Ukraine\n\nhttps://t.co/pMlyVT1kBh*
- *#Ukraine has mobilized Verka Seduchka along its eastern border*

We can clearly see that the algorithm detects the anomalies which make more sense such as coordinates and statements. From that perspective it looks like the algorithm with features work significantly better.

Normal Instances:

- *A majority of Europeans believe that Russia will invade Ukraine this year – and that Nato and the EU should stand by its eastern European ally in an armed conflict with Moscow, a continent-wide poll has found.\n\nhttps://t.co/PQ0FZLHteg*
- *Sec. of State Antony Blinken, speaking at the State Department, confirmed the U.S. had delivered a written response to Moscow security demands as Russia amassed troops on its borders with Ukraine.*
- *President Joe Biden conferred on Sunday with Ukraine's leader over the Russian troop buildup near its border, promising that the U.S. and allies will act "decisively" if Russia further invades the Eastern European nation.\n\nhttps://t.co/4z4qvqVI7M*

Looking at the normal instances it can be seen again that the model is able to detect the oficial headlines and news as normal instances which is another indicator of good performing model.

4.4.3 Unsupervised Machine Learning Algorithm K-MEANS

This section tackles to another very popular unsupervised machine learning algorithm K-means. The difference in K-means will be that the experts must themselves investigate and decide which cluster represent which class exactly. Initially, it is important to look at the results without any features to understand the difference between them. We will be using the n_clusters parameters 2 as we are assuming 2 types of classifications which are bot and non-bot. Later, we will need to analyze and decide whether there is a clear difference between the sentences classified in each cluster.

```
from sklearn.cluster import KMeans
from sklearn.feature_extraction.text import TfidfVectorizer

# Assuming your dataframe is named 'df' and contains the columns 'content_preprocessed' and 'verification_status'

# Initialize TF-IDF vectorizer
vectorizer = TfidfVectorizer()

# Apply TF-IDF vectorization on 'content_preprocessed' column
X_text = vectorizer.fit_transform(df['content_preprocessed'])

# Initialize the K-Means clustering model
model = KMeans(n_clusters=2)  # Adjust the number of clusters based on your dataset

# Assign 'Cluster 0' label to verified users
df['cluster_label'] = ['Cluster 0' if status == 'verified' else '' for status in df['verification_status']]

# Remove empty cluster labels
labeled_instances = df[df['cluster_label'] != '']
X_text_labeled = X_text[labeled_instances.index]

# Train the K-Means model on labeled instances
model.fit(X_text_labeled)

# Get the cluster labels for all instances
predicted_labels = model.predict(X_text)

# Update the cluster labels with the predicted labels
df.loc[:, 'cluster_label'] = ['Cluster 0' if label == 0 else 'Cluster 1' for label in predicted_labels]

# Print the instances and their cluster labels
for data, label in zip(df['content_preprocessed'], df['cluster_label']):
    print(f"{label}: {data}")
```

Code Snippet 6

Labels	Numbers
Cluster 0	10818
Cluster 1	5537

Table 8 : Number of Anomaly detected and Normal instance detected with K-Means with out feature

With the above code, the followings are some of the results in cluster 0 and 1:

Cluster 0:

- *@Richard62056514 @Richard75155048 @ltthompso It must be a coincidence that trump said Russia will not invade Ukraine under his watch. And here we are sitting under Biden as we spectate a bloodbath.*
- *📣 MARKET IMPACT\nUK Minister Truss: Russia's Lavrov told me has no plans to invade Ukraine\nhttps://t.co/5nPisevC01 $btc*
- *@Stonekettle There aren't any "yes or no" answers in a complex geopolitical situation.\n\nWe come to the aid of Ukraine's nextdoor neighbors, our NATO allies, which strengthens Ukraine's deterrence position. \n\nUnderstand now?*

Cluster 1:

- *Pentagon Puts 8,500 Troops On 'Heightened Alert' Over Russian Threat To Ukraine https://t.co/jilpYTnr8V via @DefenseOne*
- *CNN gets an up close look at Russian military drills in Belarus where roughly 30,000 Russian troops have been gathering along the Ukrainian border. https://t.co/yYdYqIBbF2 #belarus #lukashenko #minsk #military #osint #russia*
- *Satellite imagery shows a cluster of Russian troops and equipment near the Pripyat River in #Belarus, where a bridge appeared overnight.\nThe bridge is located about 7 km from the border with #Ukraine. https://t.co/KhxHE8N4ip*

From the above tweets of cluster 0 it is not very easy to understand the representation of the class and algorithm. Although the second tweet may look like a complete false information where seems to be an anomaly the others may look ordinary people's tweets.

According to the examples in cluster 1, some of the official news and organizations, such as the Pentagon and CNN, can be observed. Additionally, one of the tweets was validated through a Google search (NBC News, [2022]), available at [https://www.nbcnews.com/politics/national-security/defense-secretary-presents-biden-options-us-response-russia-rcna13240].

In the next step it is worthwile to add features to the model and try the model again.

```python
from sklearn.cluster import KMeans
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.preprocessing import LabelEncoder
from scipy.sparse import hstack
import pandas as pd

# Initialize TF-IDF vectorizer
vectorizer = TfidfVectorizer()

# Apply TF-IDF vectorization on 'content_preprocessed' column
X_text = vectorizer.fit_transform(df['content_preprocessed'])

# Encode the 'verification_status' column as numerical labels
label_encoder = LabelEncoder()
df['verification_status_encoded'] = label_encoder.fit_transform(df['verification_status'])

# Select the 'date_numeric' and 'verification_status_encoded' columns
X_additional = df[['verification_status_encoded', "date_numeric", "followers_count"]].values

# Combine the TF-IDF matrix and the additional columns
X_combined = hstack([X_text, X_additional])

# Initialize and train the K-Means clustering model
model = KMeans(n_clusters=2)  # Adjust the number of clusters based on your dataset
model.fit(X_combined)

# Get the cluster labels
predicted_labels = model.labels_

# Create a new column 'cluster_label' in the dataframe with the predicted cluster labels
df['kmeans_feature'] = predicted_labels

# Print the instances and their cluster labels
for data, label in zip(df['content_preprocessed'], predicted_labels):
    print(f"Cluster {label}: {data}")
```

✓ 0.4s

Code Snippet 7

When we add the date, followers and verification status as features to the model unfortunately the model underperforms with the labels:

Labels	Numbers
Cluster 0	16336
Cluster 1	19

Table 9 : Number of Anomaly detected and Normal instance detected with K-means with Features added

Taking above into consideration our observations show us the better model performance in without any features rather than with features.

5 Results and Discussion

In this section of the thesis, the primary objective is to conduct a comprehensive comparison of the results obtained from different algorithm models, with the ultimate aim of identifying the most proficient model among the ones developed. Furthermore, the inclusion of visualizations throughout this analysis will serve to enhance the readers' comprehension of the prediction distribution within the chosen model. An essential point to note initially is that, when evaluating the performance of the created models, the One-class SVM model with features stood out as the clear frontrunner. Extensive testing and analysis have consistently demonstrated its superior performance, making it a strong contender for the best-performing model in this research.

5.1 Results

The below section will cover the results obtained briefly by each 3 algorithms. By utilizing these three algorithms, we leveraged their respective strengths in identifying bot-like patterns or behaviours. One-Class SVM focuses on detecting anomalies, K-means helps in clustering similar instances, and Isolation Forest specializes in outlier detection which was the expectation from the actual algorithm. Combining their outputs and integrating domain knowledge can provide a comprehensive approach to bot detection, increasing the likelihood of accurately identifying and distinguishing bots from non-bot activity and choosing which algorithm to use for bot detection.

5.1.1 Results of Isolation Forrest

Expanding upon this, it is crucial to delve into the results generated by the Isolation Forest algorithm model. Below is the output numbers of detections by Isolation Forrest for which we described the code in practical part:

Table 10 : Result of Non Bot detected using Isolation Forrest

bot_label	Number of Occurences
Non- Bot	16355
Name: count, dtype: int64	

Upon examination, it became evident that this model severely underperformed, as it failed to detect any instances of bots within the dataset. This lack of efficacy raises concerns about the suitability of the Isolation Forest algorithm for the specific task at hand.

5.1.2 Results of K-Means

In addition to the Isolation Forest, another algorithm model that warrants exploration is the K-means algorithm. Despite showcasing a reasonably realistic distribution of clusters, it was challenging to definitively ascertain whether these clusters were indicative of bot activity or not. As such, accurately classifying the clusters proved to be a complex endeavour, casting doubt on the effectiveness of the K-means algorithm in this context. To address this, further analysis and evaluation of the K-means algorithm will be conducted to determine if there are any underlying factors influencing its inability to accurately classify the clusters. By thoroughly investigating and understanding the limitations of this model, valuable insights can be gained, enabling the development of enhanced algorithm models in future research endeavours.

Overall, the comparative analysis of the different algorithm models undertaken in this section highlights the prominent performance of the One-class SVM model with features, emphasizes the shortcomings of the Isolation Forest algorithm, and raises concerns about the efficacy of the K-means algorithm in classifying bot activity.

It is through rigorous evaluation and exploration of these models that the most effective algorithm can be identified, providing valuable guidance for future research in this field. Using the K-means algorithm, the dataset was classified in to 2 clusters; Cluster 0 (Bot) and Cluster 1(Non-bot). In the below chart the number of Non bots

based on their behaviour patterns is 5537 (33.9%) and conversely the number of Bots detected using K-means is 10,818(66.1%) users.

To remind the distribution of the model below are the pie chart and numbers that classifications detected by the algorithm.

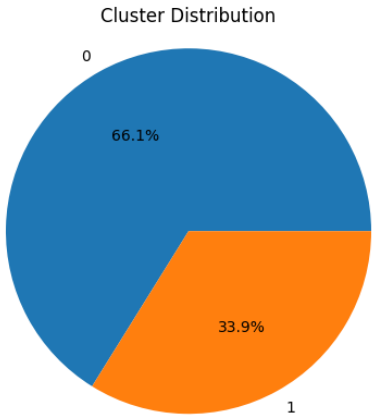

Figure 14 : Cluster distribution using K-means

kmeans_cluster_label	Number of Occurences
0	10818
1	5537
Name: count, dtype: int64	

Table 11 : Cluster distribution using K-means with out Feature added

Also the trials showed that the addition of features are not improving the model as observed in the One Class SVM.

5.1.3 Results of One-Class SVM

On the other hand, when examining the One-class SVM model, it is noteworthy that this particular algorithm demonstrated proficiency in detecting anomalies within the dataset, particularly in the context of hate speech and official organization names. Given that we have selected this algorithm as our primary model, it becomes imperative to thoroughly

analyze the distribution of the model's classification in order to gain deeper insights into its performance.

The One class SVM algorithm identfied 7,608 (46.5%) users as Anomalies (Bots), 7929(48.5%) users as Normal Instances (Unverified Users) and 818 (5%) users as Verified Users.

Below is the number of occurences resulted by the algorithm:

anomaly_label	Number of Occurences
Normal Instance (Unverified User)	7929
Anomaly Detected	7608
Normal Instance (Verified User)	818
Name: count, dtype: int64	

Table 12 : Number of occurences by One class SVM

By studying the distribution of the model's classification, we can understand how it categorizes instances and identify any patterns or trends that may emerge. This analysis will allow us to comprehend the strengths and limitations of the One-class SVM model, enabling us to make more informed decisions regarding its reliability and appropriateness for the task at hand.

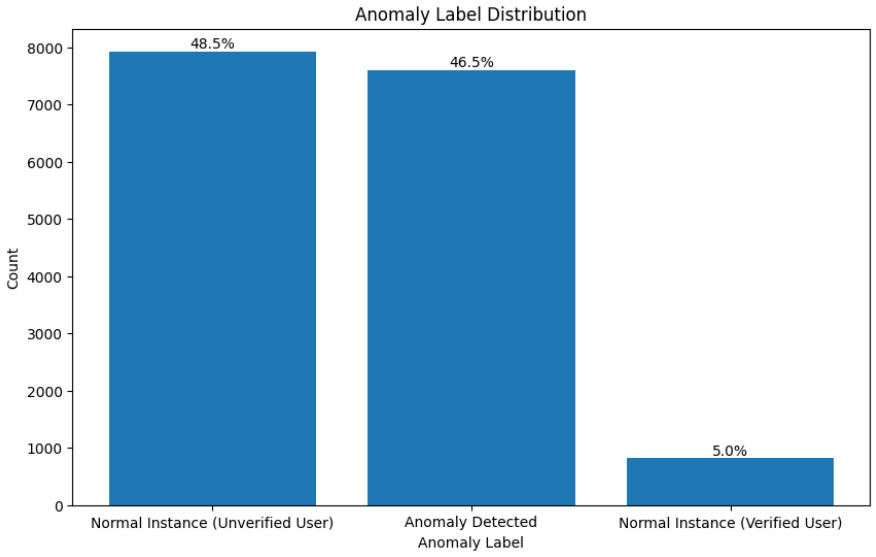

Figure 15 : Distribution of Normal and Anomalous Instances Among Unverified Users

5.2 Discussion

In this section, we present a comprehensive discussion of the results obtained from the three bot detection algorithms: Isolation Forest, K-means, and One-Class SVM. The performance of these algorithms was evaluated based on their ability to identify instances of bot activity within the dataset.

Starting with the Isolation Forest algorithm, it was observed that this model severely underperformed by failing to detect any instances of bots. The lack of efficacy raises concerns about the suitability of the Isolation Forest algorithm for the specific bot detection task at hand. Despite its potential for outlier detection, the algorithm failed to identify the anomalous patterns associated with bot behaviour in the dataset.

Moving to the K-means algorithm, it exhibited a reasonably realistic distribution of clusters. Upon analysis we also were able to see some of the tweets which could be realistically classified as bots or non-bots. Taking this into consideration we can clearly say that it was a better performing model than Isolation Forrest for our use case. However, accurately classifying these clusters as indicative of bot activity or not posed a challenge. The complexity of definitively determining the relevance of each cluster in identifying bots raised doubts about the effectiveness of the K-means algorithm in this context. Additional analysis and evaluation of the K-means algorithm are required to understand the underlying factors influencing its inability to accurately classify the clusters. Identifying and understanding these limitations will provide valuable insights for the development of enhanced algorithm models in future research.

In contrast, the One-Class SVM model demonstrated proficiency in detecting anomalies within the dataset, particularly in the context of hate speech and official organization names. The selection of One-Class SVM as the primary model proved appropriate, as it showed promising performance in identifying instances of bot activity. The thorough analysis of the model's classification distribution revealed its effectiveness in categorizing instances and recognizing patterns or trends associated with bot behaviour. These findings underscore the strengths of the One-Class SVM model in bot detection and support its reliability for the task at hand.

The addition of features in the models did not yield significant improvements, as observed in the One-Class SVM performance. However, the comparative analysis of the different algorithm models highlighted the superior performance of the One-Class SVM model with features, the underperformance of the Isolation Forest algorithm, and the limitations encountered with the K-means algorithm in accurately classifying bot activity.

Overall, our findings emphasize the importance of evaluation and exploration of different algorithm models to identify the most effective approach for bot detection. The One-Class SVM model stood out as the best-performing algorithm, demonstrating its proficiency in identifying bot-like patterns and behaviours within the dataset. Further research and investigation into the limitations of the other algorithms will provide valuable insights for future enhancements in bot detection algorithms. The results obtained from our experimentation shed light on the advantages and limitations of the Isolation Forest, K-means, and One-Class SVM algorithms in the context of bot detection. The knowledge gained through this analysis will guide future research endeavours, enabling the development of more effective and reliable algorithm models for bot detection.

5.2.1 Comparison of Pre-War and Wartime Tweets with SVM Algorithm

Before starting compare the trend of bots per months it is important to have a look on our distribution of the tweets per month in our sample. It is worthwhile to note that during the sampling, thanks to the python built in packages we sampled the data with taking into consideration the proportions of dates which enables us to say that our sample data is valid.

Below is the number of observations we have in the dataset per month:

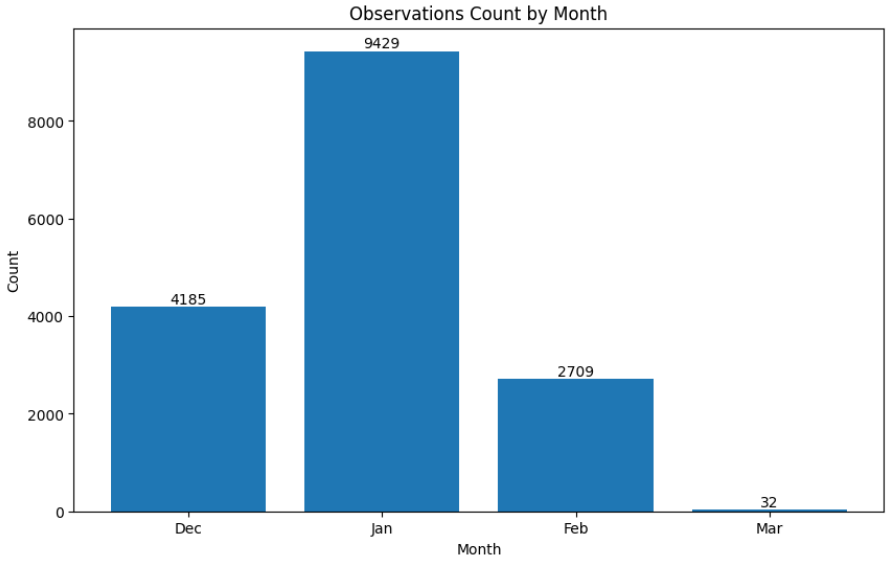

Figure 16 : Number of tweet distribution by Month

As we can see our dataset has minority in the tweets of March we will try to compare the proportions of anomalies and normal instances in each month and consider them valid until February.First of all we can start with the month of December 2021:

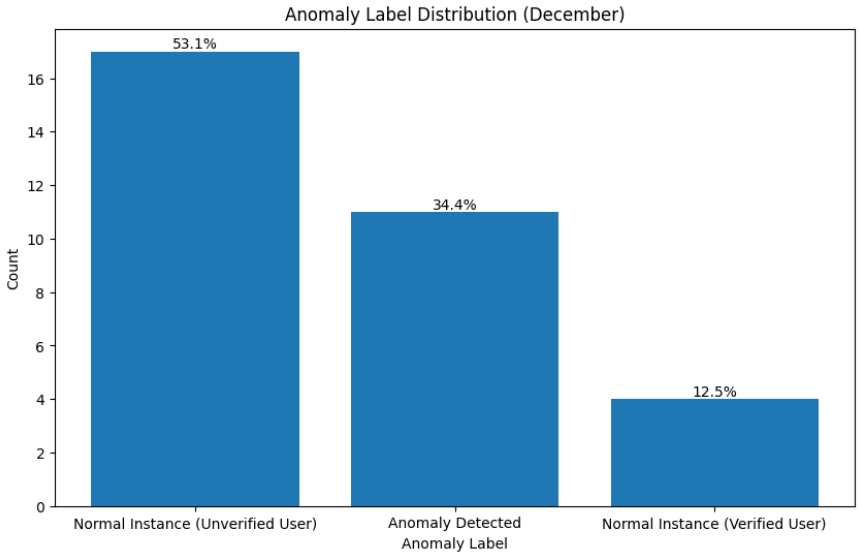

Figure 17 : Anomaly Label Distribution on the month December

As it can be seen from the above there were in total 65.6 % normal instances from which 53.1% of it were detected by the algorithm and 12.5 % was already obvious from the X (formerly known Twitter) as they were verified users.

Moving to the month of January where the tension within the war was increasing day by day following results can be seen:

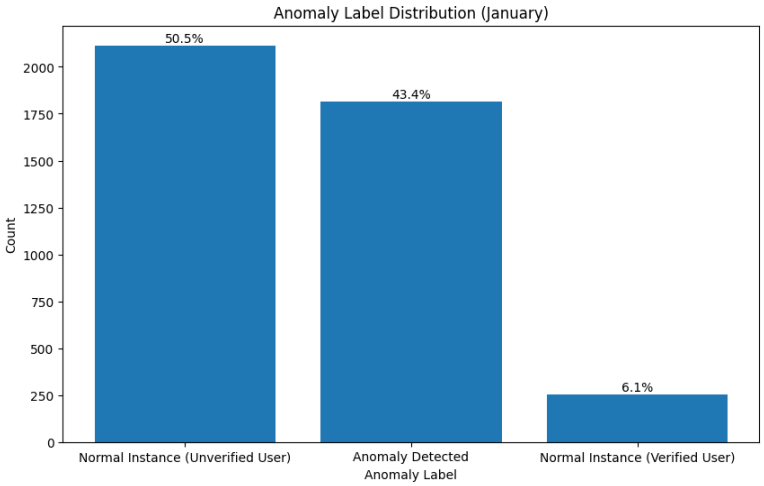

Figure 18 : Anomaly Label Distribution on the month January

Within the above chart there can be seen an increase almost by 10 percent in anomaly detections by the algorithm which is 43.4%.

Last but not least for the war time tweets as the dataset contain very small proportion of data for March, both February and March concatanated shows the following results:

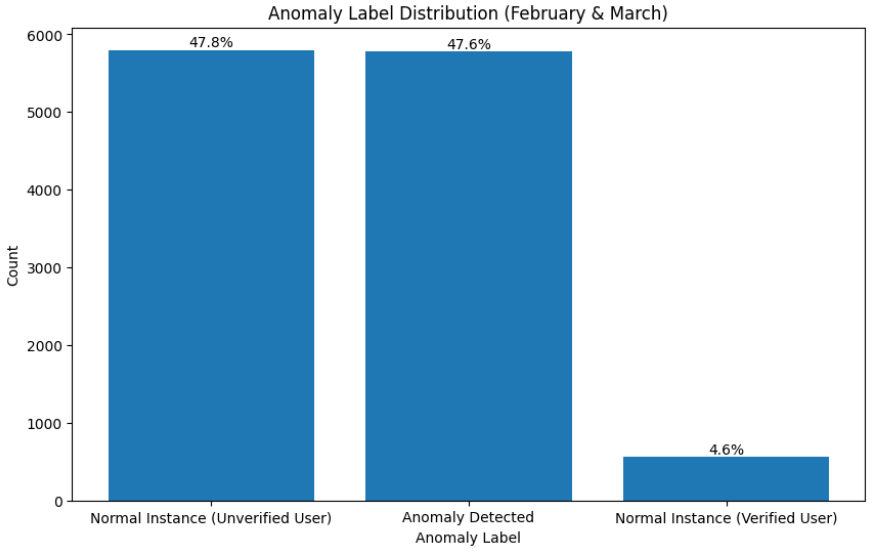

Figure 19 : Anomaly Label Distribution on the month January and March

Again there can be an increase observed by 5 % for the anomaly detection which means almost the half of the tweets in the context for Russian and Ukraine war were detected and anomalies by the machine learning algorithm.

6 Conclusion

Throughout this research, we embarked on an extensive journey to investigate and analyse a dataset comprising of collected and merged tweets. Our primary goal was to gain

insightful and valuable information from the data through descriptive statistics, deep text analysis, and the development of various machine learning (ML) algorithms.

To begin, we meticulously collected and merged the tweets, ensuring the availability of a comprehensive dataset for our analysis. This step was crucial in establishing a solid foundation for our subsequent investigations and enabled us to draw meaningful conclusions based on a wide range of data. In the initial stages of our analysis, we conducted descriptive statistics on the most significant columns of the dataset. This provided us with a clear understanding of the distribution, central tendencies, and variability of the data. By presenting these descriptive statistics, we were able to effectively summarize and communicate key characteristics and patterns within the dataset. Moving forward, we delved into deep text analysis to extract valuable insights from the textual content of the tweets. Using various techniques and methodologies, we explored the language used, sentiments expressed, and topics discussed within the tweets. Through this analysis, we gained a deeper understanding of the underlying themes and trends prevalent in the dataset.

To further our analysis, we employed a range of ML algorithms, including the Isolation Forest, K-means, and One-class SVM. By developing and implementing these algorithms, we aimed to identify patterns, detect anomalies, and classify instances within the dataset. This allowed us to assess the performance and effectiveness of each algorithm in addressing the research objectives. Upon evaluating the results of the ML algorithms, it became evident that the One-class SVM algorithm, specifically with features, outperformed the other models. This algorithm demonstrated exceptional performance in detecting anomalies, particularly in the context of hate speech and official organization names. The robustness and accuracy of the One-class SVM algorithm with features make it the preferred and recommended ML algorithm for our specific case.

In conclusion, our comprehensive analysis of the dataset, encompassing descriptive statistics, deep text analysis, and ML algorithm development, has provided us with valuable insights and conclusions. We have successfully identified the One-class SVM algorithm with features as the best-performing model for our research objectives. By proposing this algorithm as the preferred choice, we intend to guide future researchers and

practitioners in leveraging its capabilities for similar tasks. It is important to acknowledge that this research is not without limitations. While we have made significant strides in analysing the dataset and selecting an optimal ML algorithm, there may still be opportunities for further refinement and exploration. Additionally, the generalizability of our findings to other datasets or contexts should be approached with caution. Overall, the findings of this research contribute to the growing body of knowledge in the field of data analysis, text analysis, and ML algorithms. By leveraging the insights gained from this study, researchers and practitioners can enhance their understanding and application of these methodologies in various domains.

7 References

Aljabri, M., Zagrouba, R., Shaahid, A., Alnasser, F., Saleh, A., & Alomari, D. M. (2023). Machine learning-based social media bot detection: a comprehensive literature review. *Social Network Analysis and Mining*, *13*(1). https://doi.org/10.1007/s13278-022-01020-5

Allison, R. (2014). Russian "deniable" intervention in Ukraine: How and why Russia broke the rules. *International Affairs*, *90*(6), 1255–1297. https://doi.org/10.1111/1468-2346.12170

Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. *Journal of the Academy of Marketing Science*, *48*(1), 79–95. https://doi.org/10.1007/s11747-019-00695-1

Arsenault, A. (2020a). *Microtargeting, Automation, and Forgery: Disinformation in the Age of Artificial Intelligence*. 65. https://ruor.uottawa.ca/bitstream/10393/40495/1/ARSENAULT%2C Amelia_20201 - WEB.pdf

Arsenault, A. (2020b). *Microtargeting, Automation, and Forgery: Disinformation in the Age of Artificial Intelligence*. 65.

Baines, P. R., & O'Shaughnessy, N. J. (2014a). Political Marketing and Propaganda: Uses, Abuses, Misuses. *Journal of Political Marketing*, *13*(1–2), 1–18. https://doi.org/10.1080/15377857.2014.866018

Baines, P. R., & O'Shaughnessy, N. J. (2014b). Political Marketing and Propaganda: Uses, Abuses, Misuses. *Journal of Political Marketing*, *13*(1–2), 1–18. https://doi.org/10.1080/15377857.2014.866018

Binsaeed, K., Stringhini, G., & Youssef, A. E. (2020). Detecting Spam in Twitter Microblogging Services: A Novel Machine Learning Approach based on Domain Popularity. *International Journal of Advanced Computer Science and Applications*, *11*(11), 11–22. https://doi.org/10.14569/IJACSA.2020.0111103

Boichak, O., Hemsley, J., Jackson, S., Tromble, R., & Tanupabrungsun, S. (2021a). Not the Bots You Are Looking For: Patterns and Effects of Orchestrated Interventions in the U.S. and German Elections. *International Journal of Communication*, *15*, 814–839.

Boichak, O., Hemsley, J., Jackson, S., Tromble, R., & Tanupabrungsun, S. (2021b). Not the Bots You Are Looking For: Patterns and Effects of Orchestrated Interventions in the U.S. and German Elections. *International Journal of Communication*, *15*, 814–839.

Bose, R., Dey, R. K., Roy, S., & Sarddar, D. (2019). Analyzing political sentiment using Twitter data. In *Smart Innovation, Systems and Technologies* (Vol. 107). Springer Singapore. https://doi.org/10.1007/978-981-13-1747-7_41

Chen, E., & Ferrara, E. (2023). *Tweets in Time of Conflict: A Public Dataset Tracking the Twitter Discourse on the War between Ukraine and Russia*. https://github.com/echen102/ukraine-russia.

Ciotti, M., Angeletti, S., Minieri, M., Giovannetti, M., Benvenuto, D., Pascarella, S., Sagnelli, C., Bianchi, M., Bernardini, S., & Ciccozzi, M. (2020). COVID-19 Outbreak: An Overview. *Chemotherapy*, *64*(5–6), 215–223. https://doi.org/10.1159/000507423

Daya, A. A. (2019). *BotChase: Graph-Based Bot Detection Using Machine Learning*.

Dhingra, M., & Mudgal, R. K. (2019). *Historical Evolution of Social Media: An Overview*. https://ssrn.com/abstract=3395665

Fedorenko, V. L., & Fedorenko, M. V. (2022). Russia's Military Invasion of Ukraine in 2022: Aim, Reasons, and Implications. *Krytyka Prawa*, *14*(1), 7–42. https://doi.org/10.7206/kp.2080-1084.506

Garcia, M. B., & Cunanan-Yabut, A. (2022). Public Sentiment and Emotion Analyses of Twitter Data on the 2022 Russian Invasion of Ukraine. *Proceedings - 2022 9th International Conference on Information Technology, Computer and Electrical Engineering, ICITACEE 2022*, 242–247. https://doi.org/10.1109/ICITACEE55701.2022.9924136

García-López, J. C., Pinos-Rodríguez, J. M., García-Galicia, I. A., Galicia-Juárez, G. B., Gorostiola-Herrera, M. L., & Camacho-Escobar, M. A. (2011). Efecto Del Uso De Una Enzima Y Sistema De Alimentación Sobre Productividad En Pavos. *Archivos de Zootecnia*, *60*(230), 297–300. https://doi.org/10.4321/S0004-05922011000200015

Ghahramani, Z. (2004). *Unsupervised Learning* *. http://www.gatsby.ucl.ac.uk/~zoubin

Gilani, Z., Farahbakhsh, R., Tyson, G., & Crowcroft, J. (2019). A large-scale behavioural analysis of bots and humans on twitter. *ACM Transactions on the Web*, *13*(1). https://doi.org/10.1145/3298789

Hayawi, K., Saha, S., Masud, M. M., Mathew, S. S., & Kaosar, M. (2023). Social media bot detection with deep learning methods: a systematic review. In *Neural Computing and Applications* (Vol. 35, Issue 12, pp. 8903–8918). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s00521-023-08352-z

Heidari, M., Jones, J. H. J., & Uzuner, O. (2021, April 21). An empirical study of machine learning algorithms for social media bot detection. *2021 IEEE International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021 - Proceedings*. https://doi.org/10.1109/IEMTRONICS52119.2021.9422605

Jones, M. O. (2019). Propaganda, fake news, and fake trends: The weaponization of Twitter bots in the Gulf crisis. *International Journal of Communication*, *13*, 1389–1415.

Kellner, A., Wressnegger, C., & Rieck, K. (2020). What's all that noise: Analysis and detection of propaganda on Twitter. *Proceedings of the 13th European Workshop on Systems Security, EuroSec 2020*, 25–30. https://doi.org/10.1145/3380786.3391399

Khanday, A. M. U. D., Khan, Q. R., & Rabani, S. T. (2021). Identifying propaganda from online social networks during COVID-19 using machine learning techniques. *International Journal of Information Technology (Singapore)*, *13*(1), 115–122. https://doi.org/10.1007/s41870-020-00550-5

Kollanyi, B., & Howard, P. N. (2017). *Junk News and Bots during the German Federal Presidency Election: What Were German Voters Sharing Over Twitter?*

Kosmajac, D., & Keselj, V. (2019). Twitter bot detection using diversity measures. *ICNLSP 2019 - Proceedings of the 3rd International Conference on Natural Language and Speech Processing*, 1–8.

Kumar, A., Ahuja, H., Singh, N. K., Gupta, D., Khanna, A., & J. P. C. Rodrigues, J. (2018). Supported matrix factorization using distributed representations for personalised recommendations on twitter. *Computers and Electrical Engineering*, *71*, 569–577. https://doi.org/10.1016/j.compeleceng.2018.08.007

La Gatta, V., Wei, C., Luceri, L., Pierri, F., & Ferrara, E. (2023). Retrieving false claims on Twitter during the Russia-Ukraine conflict. *ACM Web Conference 2023 - Companion of the World Wide Web Conference, WWW 2023*, 1317–1323. https://doi.org/10.1145/3543873.3587571

Luo, L., Zhang, X., Yang, X., & Yang, W. (2020). Deepbot: A Deep Neural Network based approach for Detecting Twitter Bots. *IOP Conference Series: Materials Science and Engineering*, *719*(1), 8–12. https://doi.org/10.1088/1757-899X/719/1/012063

Nazarov, V. L., Gorbunov, E. V., & Kolegova, N. S. (2021). Features of Propaganda and Manipulation in the Modern Information Space of New Media. *KnE Social Sciences*, *2020*, 246–253. https://doi.org/10.18502/kss.v5i2.8358

Newsletter, C. P. (2018). *Fake News and the Politics of Misinformation*. *28*(2), Barbara, Carter and Carter, Little.

Nguyen, H. D., Nguyen, D. Q., Nguyen, C. D., To, P. T., Nguyen, D. H., Nguyen-Gia, H., Tran, L. H., Tran, A. Q., Dang-Hieu, A., Nguyen-Duc, A., & Quan, T. (2024). Supervised learning models for social bot detection: Literature review and benchmark[Formula presented]. *Expert Systems with Applications*, *238*. https://doi.org/10.1016/j.eswa.2023.122217

Ratkiewicz, J., Conover, M. D., Meiss, M., Gonçalves, B., Flammini, A., & Menczer, F. (2011). *Detecting and Tracking Political Abuse in Social Media*. www.aaai.org

Rovetta, S., Suchacka, G., & Masulli, F. (2020). Bot recognition in a Web store: An approach based on unsupervised learning. *Journal of Network and Computer Applications*, *157*. https://doi.org/10.1016/j.jnca.2020.102577

Till, C. (2021a). Propaganda through 'reflexive control' and the mediated construction of reality. *New Media and Society*, *23*(6), 1362–1378. https://doi.org/10.1177/1461444820902446

Till, C. (2021b). Propaganda through 'reflexive control' and the mediated construction of reality. *New Media and Society*, *23*(6), 1362–1378. https://doi.org/10.1177/1461444820902446

Unless, R., Act, P., Rose, W., If, T., & Rose, W. (2014). *How social media transformed pro-Russian nostalgia into violence in Ukraine*.

Vl\uaduțescu, Ștefan, & others. (2014). Communicational types of propaganda. *International Letters of Social and Humanistic Sciences*, *22*, 41–49.

Wardle, C., & Derakhshan, H. (2017). *INFORMATION DISORDER : Toward an interdisciplinary framework for research and policy making Information Disorder Toward an interdisciplinary framework for research and policymaking*. www.coe.int

Wu, J., Teng, E., & Cao, Z. (2022). Twitter Bot Detection Through Unsupervised Machine Learning. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 5833–5839. https://doi.org/10.1109/BigData55660.2022.10020983

Yaqub, U., Sharma, N., Pabreja, R., Chun, S. A., Atluri, V., & Vaidya, J. (2018, May 30). Analysis and visualization of subjectivity and polarity of twitter location data. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3209281.3209313

Zannettou, S., Sirivianos, M., Caulfield, T., Stringhini, G., De Cristofaro, E., & Blackburn, J. (2019). Disinformation warfare: Understanding state-sponsored trolls on twitter and their influence on the web. *The Web Conference 2019 - Companion of the World Wide Web Conference, WWW 2019*, 218–226. https://doi.org/10.1145/3308560.3316495

Ukraine-war-tweets-dataset-65-days. *Kaggle* [online]. 2022 [cit. 2023-03-31]. Available from: https://www.kaggle.com/datasets/foklacu/ukraine-war-tweets-dataset-65-days/data

Datacamp, (2023) *What is Tokenization?*

https://www.datacamp.com/blog/what-is-tokenization

Pulkit, Sharma, (2023). *The Ultimate Guide To K-Means Clustering: Definition, Methods and Applications*
https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/#:~:text=K%2Dmeans%20clustering%20is%20a,assigned%20cluster%20mean%20is%20minimized.

Scikit-Learn, ND, *One-Class SVM with non-linear kernel(RBF).*
https://scikit-learn.org/stable/auto_examples/svm/plot_oneclass.html#:~:text=One%2Dclass%20SVM%20is%20an,%2C%202)%20X_train%20%3D%20np.

Akshara, (2024)*, Anomaly Detection Using Isolation Forrest – A complete Guide*
https://www.analyticsvidhya.com/blog/2021/07/anomaly-detection-using-isolation-forest-a-complete-guide/

8 List of tables, figures, code snippets and abbreviations

8.1 List of tables

Table 1 : Types of propaganda(Baines & O'Shaughnessy, 2014a) 26
Table 2 : key words for German election (Kellner et al., 2020) 31
Table 3 : Summary of Missing Data 43
Table 4 : Language Distribution by Number of Occurrences 46
Table 5 : Twitter Followers Distribution by Range 47
Table 6 : Number of Anomaly detected and Normal instance detected with One class SVM with out Feature 59
Table 7 : Number of Anomaly detected and Normal instance detected with One class SVM with Features added 62
Table 8 : Number of Anomaly detected and Normal instance detected with K-Means with out feature 64
Table 9 : Number of Anomaly detected and Normal instance detected with K-means with Features added 66
Table 10 : Result of Non Bot detected using Isolation Forrest 67
Table 11 : Cluster distribution using K-means with out Feature added 69
Table 12 : Number of occurences by One class SVM 70

8.2 List of Figure

Figure 1 : Pipeline of supervised machine learning detector (Nguyen et al., 2024) 20
Figure 2 : Amplification patterns in the German election-duplicate retweeters(Boichak et al., 2021b) 32
Figure 3 : The distribution of tweet rate by follower count in the US elcetion(Boichak et al., 2021b) 33
Figure 4 : Schema of Data Analysis workflow 40
Figure 5 : Distribution of search words 44
Figure 6 : Language distribution chart 45
Figure 7 : Distribution of Verified and Unverified users 48
Figure 8 : Top 10 most frequent words 49
Figure 9 : Top 10 Word Frequencies 50
Figure 10 : Word Cloud Library 52
Figure 11 : Top 10 word frequencies in pre-processed texts 53
Figure 12 : Top 10 word frequencies in Verified Accounts 54
Figure 13 : Relationship between word frequency and sentiment 55
Figure 14 : Cluster distribution using K-means 69
Figure 15 : Distribution of Normal and Anomalous Instances Among Unverified Users .. 70
Figure 16 : Number of tweet distribution by Month 73
Figure 17 : Anomaly Label Distribution on the month December 73
Figure 18 : Anomaly Label Distribution on the month January 74
Figure 19 : Anomaly Label Distribution on the month January and March 74

8.3 List of code snippets

```python
from sklearn.cluster import KMeans
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.preprocessing import LabelEncoder
from scipy.sparse import hstack
import pandas as pd

# Initialize TF-IDF vectorizer
vectorizer = TfidfVectorizer()

# Apply TF-IDF vectorization on 'content_preprocessed' column
X_text = vectorizer.fit_transform(df['content_preprocessed'])

# Encode the 'verification_status' column as numerical labels
label_encoder = LabelEncoder()
df['verification_status_encoded'] = label_encoder.fit_transform(df['verification_status'])

# Select the 'date_numeric' and 'verification_status_encoded' columns
X_additional = df[['verification_status_encoded', "date_numeric", "followers_count"]].values

# Combine the TF-IDF matrix and the additional columns
X_combined = hstack([X_text, X_additional])

# Initialize and train the K-Means clustering model
model = KMeans(n_clusters=2)  # Adjust the number of clusters based on your dataset
model.fit(X_combined)

# Get the cluster labels
predicted_labels = model.labels_

# Create a new column 'cluster_label' in the dataframe with the predicted cluster labels
df['kmeans_feature'] = predicted_labels

# Print the instances and their cluster labels
for data, label in zip(df['content_preprocessed'], predicted_labels):
    print(f"Cluster {label}: {data}")
```

✓ 0.4s

```python
from sklearn.cluster import KMeans
from sklearn.feature_extraction.text import TfidfVectorizer

# Assuming your dataframe is named 'df' and contains the columns 'content_preprocessed' and 'verification_status'

# Initialize TF-IDF vectorizer
vectorizer = TfidfVectorizer()

# Apply TF-IDF vectorization on 'content_preprocessed' column
X_text = vectorizer.fit_transform(df['content_preprocessed'])

# Initialize the K-Means clustering model
model = KMeans(n_clusters=2)  # Adjust the number of clusters based on your dataset

# Assign 'Cluster 0' label to verified users
df['cluster_label'] = ['Cluster 0' if status == 'verified' else '' for status in df['verification_status']]

# Remove empty cluster labels
labeled_instances = df[df['cluster_label'] != '']
X_text_labeled = X_text[labeled_instances.index]

# Train the K-Means model on labeled instances
model.fit(X_text_labeled)

# Get the cluster labels for all instances
predicted_labels = model.predict(X_text)

# Update the cluster labels with the predicted labels
df.loc[:, 'cluster_label'] = ['Cluster 0' if label == 0 else 'Cluster 1' for label in predicted_labels]

# Print the instances and their cluster labels
for data, label in zip(df['content_preprocessed'], df['cluster_label']):
    print(f"{label}: {data}")
```

```python
from sklearn import svm
from sklearn.feature_extraction.text import TfidfVectorizer
import pandas as pd

# Convert preprocessed text data into numerical feature vectors
tfidf_vectorizer = TfidfVectorizer()
X = tfidf_vectorizer.fit_transform(df['content_preprocessed'])

# Train the One-Class SVM model
ocsvm_model = svm.OneClassSVM()
ocsvm_model.fit(X)

# Make predictions on the same feature vectors
predictions = ocsvm_model.predict(X)

# Map the numerical predictions to labels
labels = ['Normal Instance (Unverified User)' if p == 1 else 'Anomaly Detected' for p in predictions]

# Update the labels based on the verification status
for index, row in df.iterrows():
    if row['verification_status'] == 'verified':
        labels[index] = 'Normal Instance (Verified User)'

# Create a new column 'anomaly_label' in the DataFrame with the predicted anomaly labels
df['anomaly_label'] = labels

# Print the instances and their anomaly labels
for index, row in df.iterrows():
    print(f"Instance {index + 1}: {row['content_preprocessed']} | Anomaly Label: {row['anomaly_label']}")
```

✓ 59.6s


```python
from sklearn import svm
from sklearn.feature_extraction.text import TfidfVectorizer
import pandas as pd

# Convert preprocessed text data into numerical feature vectors
tfidf_vectorizer = TfidfVectorizer()
X = tfidf_vectorizer.fit_transform(df['content_preprocessed'])

# Train the One-Class SVM model
ocsvm_model = svm.OneClassSVM()
ocsvm_model.fit(X)

# Make predictions on the same feature vectors
predictions = ocsvm_model.predict(X)

# Map the numerical predictions to labels
labels = ['Normal Instance (Unverified User)' if p == 1 else 'Anomaly Detected' for p in predictions]

# Update the labels based on the verification status
for index, row in df.iterrows():
    if row['verification_status'] == 'verified':
        labels[index] = 'Normal Instance (Verified User)'

# Create a new column 'anomaly_label' in the DataFrame with the predicted anomaly labels
df['anomaly_label'] = labels

# Print the instances and their anomaly labels
for index, row in df.iterrows():
    print(f"Instance {index + 1}: {row['content_preprocessed']} | Anomaly Label: {row['anomaly_label']}")
```

✓ 59.6s

```python
from sklearn.ensemble import IsolationForest
from sklearn.feature_extraction.text import TfidfVectorizer
import pandas as pd

# Convert preprocessed text data into numerical feature vectors
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(df['content_preprocessed'])

# Train the Isolation Forest model
isolation_forest = IsolationForest()
isolation_forest.fit(X)

# Predict the anomaly scores
anomaly_scores = isolation_forest.decision_function(X)

# Set a threshold to classify instances as bot or non-bot
threshold = -0.5  # Adjust the threshold based on your data and requirements

# Create a new column 'bot_label' in the DataFrame with the predicted labels
df['bot_label'] = ['Bot' if score < threshold else 'Non-Bot' for score in anomaly_scores]

# Print the instances and their bot labels
for index, row in df.iterrows():
    print(f"Instance {index + 1}: {row['content_preprocessed']} | Bot Label: {row['bot_label']}")
```

```python
import spacy
from spacy.lang.en import English
from nltk.corpus import stopwords
import re
import pandas as pd

# Load the English language model in spaCy
nlp = spacy.load('en_core_web_sm', disable=['parser', 'ner'])

# Download stopwords (optional)
stop_words = set(stopwords.words('english'))

# Custom words to remove
custom_words = ['Ukraine', 'Russia', 'Russian', 'war', 'US', 'invade', 'would', 'amp', 'StandWithUkraine', 'invasion',
                'country', 'Ukrainian', 'like', 'want', 'going', 'Europe', 'think', 'world', 'says', 'said', 'one',
                'countries', 'get', 'right', 'back', 'know', 'go', 'time', 'President', 'near', 'support', 'need', 'even', 'take']

# Function to remove stopwords, special characters, and lemmatization using spaCy
def preprocess_text(text):
    # Tokenize the text
    doc = nlp(text)

    # Remove stopwords, special characters, and custom words
    filtered_tokens = [re.sub(r'[^\w\s]', '', token.lemma_.lower()) for token in doc if token.lemma_.lower() not in stop_words.union(custom_words)]
    filtered_tokens = [token for token in filtered_tokens if len(token) > 1]  # Remove individual characters

    return ' '.join(filtered_tokens)

# Apply the preprocess_text function to the 'content_cleaned' column
df_english['content_preprocessed'] = df_english['content_cleaned'].apply(preprocess_text)

# Print the updated DataFrame
print(df_english)
```

56m 15.1s

8.4 List of abbreviations

CIA - Central Intelligence Agency

CNN - Convolutional Neural Network

DL - Deep Learning

GAN - Generative Adversarial Network

GCC - Gulf Cooperation Council

GNN - Graph Neural Network

KNN - K-Nearest Neighbors

LSTMS - Long Short-Term Memory Systems

ML - Machine Learning

MOSSAD - (The national intelligence agency of Israel)

NATO - North Atlantic Treaty Organization

NLP - Natural Language Processing

NLTK - Natural Language Toolkit

PCA - Principal Component Analysis

RNN - Recurrent Neural Network

SVM - Support Vector Machine

UAE - United Arab Emirates

9 Appendix

Source Code
Appendix.pdf

PDF form which contains all the code, different models, and graphs