Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Information Technology (FEM)

Bachelor Thesis

Fake profile detection on social networks

Kishor Kanji Kerai

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

BACHELOR THESIS ASSIGNMENT

Bc. Kishor Kanji Kerai

Informatics

Thesis title

Fake profile detection on social networks

Objectives of thesis

The main aim of the research is the analysis of fake profile detection, reasons for rising malicious activities and providing effective methods to protect ourselves from such things.

Secondary objectives include:

- Formalise the difference between bots and humans
- Compare the detection measures to resolve the issues related to fake accounts
- Create a critical review of previous papers work related to fake profile detection
- Creating a conclusion and practising all the combined measures to protect users from fake accounts and bots.

Methodology

In theoretical parts, methods of induction and deduction will be used, finding statics and data from famous blogs, books, and by the grace of the author's work. In the practical part, Machine learning has been discussed. The practical part gives all the ideas on how to manage these activities and be aware.

Data collection methods: All data required to complete this thesis will be taken from research, induction and deduction, statics and comparative methods selected according to available data sources.

The proposed extent of the thesis

40-50

Keywords

Social media analysis, Security and Privacy, Fake profile Detection, Data mining and techniques, survey.

Recommended information sources

C. Xiao, D. M. Freeman, and T. Hwa, ''Detecting clusters of fake accounts in online social networks,'' in Proc. 8th ACM Workshop Artif. Intell. Secur., 2015, pp.

J. P. Dickerson, V. Kagan, and V. S. Subrahmanian, ''Using sentiment to detect bots on Twitter: Are humans more opinionated than bots?'' in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2014, pp. 620–627.

S. Gurajala, J. S. White, B. Hudson, B. R. Voter, and J. N. Matthews, ''Profile characteristics of fake Twitter accounts,'' Big Data Soc., vol. 3, no. 2, p. 2053951716674236, 2016, doi: 10.1177/2053951716674236.

T. Tuna et al., ''User characterization for online social networks,'' Social Netw. Anal. Mining, vol. 6, no. 1, p. 104, 2016.

Y. Li, O. Martinez, X. Chen, Y. Li, and J. E. Hopcroft, ''In a world that counts: Clustering and detecting fake social engagement at scale,'' in Proc. 25th Int. Conf. World Wide Web, 2016, pp. 111–120.

Expected date of thesis defence

2022/23 WS – FEM

The Bachelor Thesis Supervisor

Ing. Tomáš Vokoun

Supervising department

Department of Information Technologies

Electronic approval: 23. 8. 2021

doc. Ing. Jiří Vaněk, Ph.D.

Head of department

Electronic approval: 5. 10. 2021

Ing. Martin Pelikán, Ph.D.

Dean

Prague on 13. 03. 2023

Declaration

I declare that I have worked on my bachelor thesis titled "Fake profile detection on social networks" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the bachelor thesis, I declare that the thesis does not break any copyrights.

In Prague on 2024 _____

Acknowledgement

I would like to thank The **Czech University of Life Sciences Prague** and my supervisor, **Ing. Tomas Vokoun**, allowing me to conduct research on this excellent thesis topic. They were constantly encouraging while I was researching on my thesis.

I'd also like to thank all my family, friends and other persons, for their advice and support during my work on this thesis.

Fake profile detection on social networks

Abstract

Fake profiles are becoming more and more common due to the social networking sites' explosive growth, which puts users' security, privacy, and reliability of online interactions at serious risk. This thesis suggests a thorough examination of cutting-edge techniques to enhance the identification of phoney social network profiles. The study will analyse prior research that covers a variety of conventional and cutting-edge methodologies, such as network analysis techniques, machine learning algorithms, data mining strategies, and natural language processing techniques. The study will also investigate a number of variables, including profile parameters, activity patterns, social interactions, and content semantics, that are employed in the detection of phoney profiles. The goal of the research is to develop creative solutions that will improve accuracy and efficiency while addressing the shortcomings and difficulties of the detection methods used today. This will entail developing innovative frameworks and algorithms that make use of cutting-edge technology including sentiment analysis, graph neural networks, and deep learning.

Keywords: Social media analysis, Security and Privacy, Fake profile Detection, Data mining and techniques, survey.

Detekce falešného profilu na sociálních sítích

Abstraktní

Falešné profily se stále více běžnější díky výbušnému růstu sociálních sítí, což zvyšuje bezpečnost uživatelů, soukromí a spolehlivost online interakcí s vážným rizikem. Tato práce navrhuje důkladné zkoumání špičkových technik pro zvýšení identifikace falešných profilů sociálních sítí. Studie bude analyzovat předchozí výzkum, který zahrnuje řadu konvenčních a špičkových metodik, jako jsou techniky síťové analýzy, algoritmy strojového učení, strategie těžby dat a techniky zpracování přirozeného jazyka. Studie také prozkoumá řadu proměnných, včetně parametrů profilu, vzorců aktivity, sociálních interakcí a sémantiky obsahu, které se používají při detekci falešných profilů. Cílem výzkumu je vyvinout kreativní řešení, která zlepší přesnost a efektivitu a přitom se zabývají nedostatky a obtížemi detekčních metod používaných dnes. To bude zahrnovat vývoj inovativních rámců a algoritmů, které využívají špičkovou technologii, včetně analýzy sentimentu, grafových neuronových sítí a hlubokého učení.

Klíčová slova: Analýza sociálních médií, Bezpečnost a soukromí, Detekce falešných profilů, Data mining a techniky, průzkum.

Table of Contents

Table of Figures

Table of Tables

1. Introduction

Over the last few years, online social networking sites have emerged on a large scale. A large amount of data is circulated all over the web through the social media networks because social media provides a cheap and efficient way of sharing any kind of information. Social media is useful in providing a platform to many small businesses, promotions, small pages, efficient information, etc. Due to this large amount of data available, social media gains the attention of many spammers, frauds, scammers, fake people, etc. This gives birth to fake profiles, sybils (massive fake accounts), and a lot of malicious activities. People use these fake accounts to perform malicious activities, abuse people or system and to commit frauds. This leads to highly dangerous issues like privacy and security concerns, sexual harassment, impersonation and hacking, etc. To protect ourselves from all these activities, fake profile detection methods come into play and it's very important to us to be aware of these detections. The fake profiles are not always run by any person, sometimes it automatically runs through bots, which can mimic human aspects. I have Machine Learning techniques and distance measure algorithms which helps in activities like fake profile detection and confirming spam users. Fake profiles can also be identified manually in some cases by analysing the graphs and algorithm of content posted (Adebowale, M.A. et al., 2023).

The role of fake profile detection in history and future are discussed in this paper. The importance of" Fake profile detection" is devoted and confirmed with all the authors mentioned below under references. The current statistics and ways of detection are discussed in detail below (Hao, P. and Wang et al., 2019).

Online Social Networks

A collection of nodes (people, actors, organizations, countries, and states, etc.) connected by a network of links (relationships, interactions, distances, hyperlinks, etc.) is known as an online social network (OSN). Web applications that are primarily used on OSNs are developed to encourage user involvement, teamwork, and content sharing. OSNs have altered how people

think, communicate, and interact with others. Several social networking websites, including Facebook, Twitter, Flickr, People utilize websites like LinkedIn, ResearchGate, and others to conduct their social and professional endeavours. OSNs' structure is similar to that of real-world systems, therefore Communities are quite valuable since they contain a ton of user contentsignificant to academics and experts in a number of other fields, such as marketing, sociology, government, etc. To develop tactics for viral marketing, marketing firms research OSNs (Fire, M., Goldschmidt, R. and Elovici, Y., 2014).

Sociologists utilize them to study human behaviour and help businesses reach their potential customers.

The collections of nodes (people, actors, organizations, countries, and states, etc.) that make up an online social network (OSN) are joined by a network of links (relationships, interactions, distances, hyperlinks, etc.). Web applications that are predominantly OSNs are referred to as aimed for promoting user connection, teamwork, and content sharing. Onstage altered how people communicate, think, and interact with the outside world. There are numerous social networking services available now, including Facebook, Twitter, Flickr, People use websites like LinkedIn, ResearchGate, and others for social and professional pursuits Given that OSNs' structure resembles that of real-world systems, Communities are highly valued since they contain a significant volume of user content vital to the academics and a number of other academic fields like marketing, sociology, etc., politics To create viral marketing techniques, marketing firms research OSNs. Sociologists use them to examine how businesses may attract new clients and Politicians make use of them to strengthen their electoral campaigns (Al-Qurishi, M. et al., 2018).

There is a huge diversity of OSNs based on the traits. Social networks, for instance, assist users in creating online social ties with friends and family. These OSNs include Twitter, Facebook, Myspace, and others.

The most popular social network on the internet, Facebook gives members a place to connect and share information with those they know. Twitter allows users to share their ideas, viewpoints, and proposals while also getting updates from other users who are connected to them. Some social networking sites, like YouTube and Flickr, are made particularly to offer a

simple and handy way to exchange movies and photographs. There are also online networking tools like LinkedIn that are primarily created to foster users' professional development. It is an online professional network that gives users a potent way to interact with others doing the same type of work (Foody, M. et al., 2015).

Fake profile detection

I can check fake profiles by keeping a check of profiles with no profile image or profile name. This method is used to detect all fake Twitter profiles. Here, fake profiles detection is based on rules that can differentiate fake profiles from genuine or real ones. The geo-enabled field feature will come out as false since they do not want to expose their personal location in any tweets. Fake profiles either make a lot of tweets or don't make any tweets. The rules gets applied on the profile initially and, for each matching rule, a counter gets incremented, often it happens that the counter value comes out to be greater than predefined threshold, and in that case, the profile is termed as fake account (Sowmya, P. and Chatterjee, M., 2020, July).

Clone profile detection

In this module, detection clones are based on Attributes and Network similarity. I can take user profiles as input. This module follows a search mechanism in which profiles having attributes matching to that of the user's profile are looked. It calculates similarity index based on the columns obtained and if the similarity index comes out greater than the threshold, then the profile can be stated as clone, else considered normal.

Attribute similarity: Attribute similarity is analysed on the basis of similarity of attribute data or values obtained between the profiles. The attributes which are taken into consideration for similarity measures are Name, Profile Name, Language, Location and Time zone, Contact details, E-mail ids, etc. Two major similarity measures are being

followed to find the similarity between the attributes or columns. These two measures include Cosine similarity and Lowenstein distance. The Cosine similarity is used to find similarity between the words. The next measure, that is Lowenstein distance, is applied to find similarity between two sequences. The cosine similarity measure is given by.

Two different vectors can have a cosine similarity of 1 if they have the same orientation. Similarly, two vectors will have a similarity of 0 if they are fixed at 90° and -1 if they are diametrically opposed. The Lowenstein distance method is a similarity measuring system to find similarity among two sequences. If i provided with two sequences, the Lowenstein distance between them will be the minimum number of inserts, substitution or delete operations needed to convert one sequence into another.

Network similarity: Network similarity is analysed on the basis of network relationships. Here, network attributes are used such as Follower's id's attribute have been in play to find the network similarity between different profiles. Attribute Follower's id's keeps a record of the list of accounts that follows the concerned user. All clone profiles that exist try to keep their profiles as much similar as possible to the legit original account.

Since users and tweets maintain a 1-to-N relationship, I need to figure out an appropriate way for machine learning to decode the tweet documents in a classified table containing accounts.

Data acquisition involves obtaining the dataset through requests and email correspondence. Researchers gathered data from various sources including fake follower datasets, fake websites, and monitoring spam bot activity during the mayoral election in Rome, Italy. Data cleaning is then performed to standardize the dataset, converting text to lowercase, removing non-UTF characters, white spaces, special characters, and hyperlinks. This ensures uniformity and consistency in the dataset. Feature extraction is conducted using Scikit-Learn's TfidfVectorizer and Count Vectorizer to extract Bigram and tf-idf features from the transformed data. Additionally, Word2Vec features are utilized, employing a pre-trained model generated from English tweets collected via the Twitter Streaming API. This model encompasses a broad vocabulary of English words and includes a tokenizer application designed to capture the informal nature of Twitter language.

2. Objectives and Methodology

The objectives and methodology of the thesis is as follows.

2.1 Objectives

The main aim of the research is the analysis of fake profile detection, reasons for rising malicious activities and providing effective methods to protect ourselves from such things.

Secondary objectives include:

- Formalise the difference between bots and humans

- Compare the detection measures to resolve the issues related to fake accounts

- Create a critical review of previous papers work related to fake profile detection

- Creating a conclusion and practicing all the combined measures to protect users from fake accounts and bots.

2.2 Methodology

In theoretical parts, methods of induction and deduction will be used, finding statics and data from famous blogs, books, and by the grace of the author's work. In the practical part, Machine learning has been discussed. The practical part gives all the ideas on how to manage these activities and be aware.

Data collection methods: All data required to complete this thesis will be taken from research, induction and deduction, statics and comparative methods selected according to available data sources.

3. Literature review

A basic analysis and union of past investigations and scholastic distributions relevant to a particular subject or research issue comprise a writing survey. It offers a careful handle of the condition of information in the subject the way things are at the present time. A writing survey tracks down holes, inconsistencies, and subjects for more examination by assessing and integrating prior discoveries. It gives foundation, hypothetical systems, and experiences into the procedure utilized in before studies, which lays the basis for additional review.

3.1 Overview

OSNs have changed how individuals convey, think, and communicate with others. People utilize a plenty of social systems administration locales nowadays, including Facebook, Twitter, Flickr, LinkedIn, ResearchGate, and some more, for both social and professional purposes. OSNs are very important on the grounds that they incorporate an abundance of client content that is vital for researchers and professionals in a few different areas, similar to legislative issues, social science, showcasing, etc. This is on the grounds that their construction is like that of genuine networks. Sociologists use OSNs to inspect human way of behaving, advertisers use them to make viral showcasing procedures and contact new crowds, and lawmakers use them to support their political missions. The huge measure of data contained in these OSNS with respect to a client's social, personal, and professional life has not drawn a lot of consideration from specialists or cybercriminals. These web-based criminals get sufficiently close to OSNs by making fake personalities or via completing different identity burglary strategies on dynamic clients, like cloning, mocking, and different techniques, to acquire their accreditations. These programmers, especially the talented assailants, likewise make different bots to manage their made-up personas with next to zero human contribution (Adewole, K.S. et al., 2020).

On social media destinations like Facebook and Twitter, a rising number of programmers are making bogus characters with an end goal to get to clients' personal and social media information, to promote a specific organization or individual, to criticize a client, and so on.

Foes might utilize professional sites like LinkedIn and ResearchGate to follow conduct and power clients to uncover personal data with an end goal to acquire the trust of different clients or business professionals. They likewise target dating administrations fully intent on framing private or sexual connections, or to get gifts, cash, or different advantages. A review inspected the different protection and security worries that OSN clients experience and gave a reasonable arrangement of thoughts to guard clients both on the web and disconnected.

While certain individuals make numerous personalities for amusement purposes or to interface with specific companions, and so on, fake profiles aren't dependably harmful. However, in light of the fact that they disregard the guidelines and laws of the help, they are seen as unlawful. Coming up next are cases of rules and guidelines as they apply to OSN. It shouldn't utilize computerized instruments like bots and bugs to get to the organization, send any unsafe or unlawful substance, or gather client information. It isn't fitting for the proprietor to have different personal accounts. Facebook claims that any record that is kept up by somebody other than their principal account is false. A huge number of counterfeit profiles can be found on notable social media locales like Facebook and Twitter, particularly in the Chinese and Indian business sectors. Social systems administration specialist co-ops utilize different procedures to guarantee the security of their clients (Arshad, H. et al., 2020).

Fake Profile Attributes

The main prerequisite for identifying phoney profiles on social networking sites is to examine the traits that will set them apart from genuine profiles. A suitable and useful feature set must be prepared in order to construct an effective false profile detector. The traits can be manually examined on social networking sites or investigated via a literature review. It's also likely that some of the traits described in literature won't work today because adversaries continue to alter their behaviour in order to trick and get past detection systems (Rout, R.R. et al., 2020).

Several academics have periodically recognised several characteristics of online profiles to train their fake profile detection programmes.

Network-Based Attributes

Like individuals do, in actuality, when they cooperate with companions, make new companions, and talk about issues, OSN clients partake in everyday social exercises like communicating with online companions, making new companions, and joining new networks (gatherings, pages, and so forth.). They likewise fabricate an organization of trust among companions. These elements, which incorporate client made gatherings, the amount of companion demands acknowledged (in degree), the amount of companion demands sent (outdegree), how much a hub fills in as a scaffold between different hubs (betweenness centrality), the hub that is nearest to each and every hub in the organization (closeness centrality, etc, are known as organization credits.

In view of these organization characteristics, various scholastics made models for the distinguishing proof of fake profiles on OSNs. Utilizing network properties to distinguish deceitful profiles on the Twitter network is recommended. Three angles have been recognized as obvious profiles: genuine social cooperations, the advancement of OSN companions after some time, and changes in the OSN organization's construction over the long run. The third trademark to distinguish sham records checks out at the typical level of hubs and the quantity of singleton companions. Various clients have various characteristics relying upon the kind of social organization they join.

Content Based Attributes

The text, images, videos, and other media that a person posts or shares on their profile are referred to as content. Some examples of content include the number of tags and words used in a post. The material reveals a lot about the user's habits. A user's content on the network reflects both his or her conduct and personality in general. Researchers have used a number of content-based features in the literature to distinguish between different types of spammers on various social networks. For the purpose of identifying sock puppet accounts on Wikipedia, the authors of employed content-based criteria like quotation length, punctuation, and usage of capital or lowercase letters. Utilizing highlights like the amount of hash labels and URLs in the message, one more concentrate on the Chinese social organization Sina Weibo had the option to distinguish spammers on the platform (Latah, M., 2020). On the social media stage Facebook,

content-based components like the amount of photographs somebody has transferred and how much pictures they have been labeled in have additionally been utilized to distinguish fake profiles. The message title is yet another feature. Twitter's "hash tag" feature allows us to specify the topic of a specific post. The users typically talk about a certain set of interests they have, such as their favourite sports, movies, political parties, etc. On the other hand, phoney user-generated postings are frequently unconnected because individuals typically post about their preferred topics, which makes them largely related to one another. This peculiar behaviour can also be used to identify an atypical user (Rodríguez-Ruiz, J. et al., 2020).

Temporal Features

As the name suggests, temporal features are traits that are time-related, such as the time an account was created, the date and time the user last logged in, the interval between two status updates, the duration of an account's active status, etc. Creators guarantee that clients might be dubious assuming they change their announcements or do some other activity on their profiles at odd hours (during regular resting hours). Moreover, it has been noticed that social bots, which are heavily influenced by a solitary foe, can all the while enact a botnet of records, perform activities (frequently terrible ones), and afterward log out simultaneously. Scientists can utilize this time sensitive way of behaving to distinguish a social botnet (Wanda, P. and Jie, H.J., 2020).

Profile Based features

Profile components on an OSN incorporate fundamental data about a client's identity, similar to orientation, age, area, telephone number, email address, nationality, name, photograph, number of companions, work, and education (Boshmaf, Y. et al., 2016). Various investigations have utilized highlights from client profiles to recognize typical and strange individuals on various social networks. The creators of utilized many AI methods to profile-based qualities and companion data, (for example, the quantity of companions, number of follows, and so on) to distinguish spammers on the MySpace and Twitter networks (Muñoz, S.D. and Pinto, E.P.G., 2020). To recognize social bots on Twitter, an AI approach has been carried out. The creator has

used profile credits including the client profile's supporter to-devotee proportion. Additionally, in consistence with Twitter's spam strategy, on the off chance that you attempt to follow a bigger number of clients than are allowed or on the other hand in the event that the quantity of individuals you follow is not exactly the all-out number of individuals you follow, your identity should be visible as suspicious. Similarly, in a review, the creators took a gander at the special characteristics of spammers on Sina Weibo and Tencent Weibo, two Chinese microblogging networks. They did this by analyzing profile information like following/supporter proportion and record age, in addition to other things. Aside from the recently recorded qualities, there exist various other profile attributes that can be utilized to recognize peculiarities across a scope of OSNs. For instance, a client could show that they communicate in a few dialects in their "dialects known" segment, however at that point submit posts in a language that isn't broadly utilized (Quinlan, J.R., 1987).

Action Based Features

A quality that is activity based is one not entirely set in stone by the client's activities on a social organization, such posting, transferring, sharing, etc. The manner in which an individual responds to or remarks on their companions' posts on the organization likewise mirrors their way of behaving there. Since activity based characteristics, now and again alluded to as action highlights, include hostile and discriminating public way of behaving, they are fundamental for recognizing spam accounts. A concentrate on the Chinese social organization Sina Weibo claims that actions, for example, the amount of messages conveyed day to day, the amount of remarks left, the amount of preferences a post gets, and so on can be utilized to distinguish spammers on the stage. Spammers post messages multiple times more as often as possible than non-spammers, as indicated by the review's creators. To detect odd client conduct on the Facebook organization, the writers of utilized the pace of like action, a social metric that measures how much of the time an individual preferences pages. At the point when a client prefers a page a great deal in a short measure of time, this could be reason to worry.

3.2 Types of fake profiles in online social networks

On the basis of characteristics of fake profiles, I have divided them into five categories, viz., compromised profiles, cloned profiles, Sybil accounts, sock puppets, and fake bot profiles. Each category has been described individually in the following sub-sections. These categories can be considered as the different ways by which the adversaries achieve their ill aims on different online social networking platforms. These categories can be thought of as the many means by which the enemies accomplish their evil objectives on various online social networking platforms (statista, "Number et al., 2020).

3.2.1 Compromised Profiles

Compromised accounts are the real accounts but their owners don't have complete control over them, or they have lost the control to a phished or any malware agent.

As per the Face book terms and conditions, any legitimate account that is accessed by the person who is not the authorized owner of the account is considered compromised. According to authors in compromised accounts are the most difficult type of accounts to be detected as the real owner has already maintained a level of trust on the networks. Another recent study reported that more than 97% profiles are compromised rather than fake out of the total identified malicious accounts which were used to spread spam. Adversaries usually create fake profiles to steal the credentials from the real users, and once the goal is achieved, these adversaries abandon or deactivate the fake ones and start using compromised ones for the illicit activities.

Because compromised profiles have already gained some trust from others in their network, they are very valuable and difficult for service providers to identify and remove from the social network. The study reveals that the compromised real profiles spread more malicious content than other types of fake profiles. Face book assists its users to recover hacked and compromised accounts once reported. There are options such as my-account-was-hacked10 or My-Account-is Compromised on the Facebook help page, using which the users can report their compromised accounts. Usually, cyber criminals launch various phishing attacks to obtain credentials of a real

account in order to perform several unlawful activities, and this is considered a serious cybercrime. There are several reasons for a profile to get compromised such as weak passwords, virus infections, sharing passwords, etc. Users should take proper care while using social media accounts to secure their personal and social data from cybercriminals (Wani, M.A., Jabin, S., Yazdani, G. and Ahmadd, N., 2018).

3.2.2 Cloned Profiles

Profile cloning is a technique in which the adversary establishes another profile using information such as name, age, gender, profile picture, etc. of any existing real profile. In other words, I can say profile cloning is the process of stealing the victim's information in order to create one more profile to spread spam, obtain private information about the victim and the victim's associates, or engage in additional fraudulent activities like luring, slander, etc. After creating a clone profile, the cloner can begin sending scam messages under the victim's name and friend requests to others on the victim's friend list. Even identity theft can be carried out by a cunning cloner who tricks the victim's pals into disclosing a lot of personal and financial data. This is called Identity Clone Attacks (ICAs). There are two types of profile cloning attacks namely single site profile cloning and cross-site profile cloning. The attackers are typically well-funded, knowledgeable individuals with access to nearly everything, and they are in charge of accounts that have been compromised or infected. The adversary can be a strange person, but statistics show that adversary has the knowledge of victim and can be one of the victim's relatives, friend or colleague (Wani, M.A., Jabin, S., Yazdani, G. and Ahmadd, N., 2018).

A recent study made recommendations for OSN sites to enhance security and user self-defense, as well as various strategies to deal with cloning attempts. Authors in a study presented two automated ICAs namely 'profile cloning' and 'cross-site profile cloning' and proposed prototype attack system (iCLONER) to attack the five most popular OSNs including XING, StudVZ, MeinVZ, Facebook and LinkedIn. This study demonstrated that ICA techniques are very successful and that consumers do not see their enemies with much distrust. One major problem with online social networks is profile cloning. Normal users are not aware that their identities are being copied and used as a weapon to destroy their kingdom by dodgy characters. These

criminals copy all the content from victim's profile including profile name, profile picture, education, work even status updates to give it exactly the same look as the real account and exploit it to perform other cybercrimes (Bhumiratana, B., 2011).

3.2.3 Sock puppets

A sock puppet is an account created on social networking sites, blogs, discussion forums, etc. with the intention of misleading people or promoting someone or something. Put differently, sock puppets are online personas designed to deceive users in a variety of ways, such as convincing them that a specific product is a suitable investment or that a plan with a high rate of return has little risk. When it comes to OSN websites, barred users typically create new accounts—a practice known as "sock puppetry"—in order to gain access. According to the authors in, if there exist two different accounts on any news blog, social network or any discussion forum that belong to the same person, it is called a sock puppet pair. Sock puppets are created for several reasons including business promotions, fake reviews on books and movies, false campaigning, defend or support a person or an organization, etc. In case of discussion forums, sock puppets are used to engage people by deceiving others or manipulating discussions. The authors in studied the sock puppetry and showed that sock puppets have different posting behaviour than normal users in various discussion forums. In OSNs like Facebook, Twitter, etc., sock puppets are created to make more followers, false likes and also to conduct mass propaganda through retweets and comments. Establishing Sock puppets is considered as one of the main ways of online deception. Nowadays, sock puppets are used for false marketing- an example of astroturfing, in which the people artificially stimulate online conversation and positive reviews about a particular product, brand or service. Sock puppets are frequently used on social media sites to draw attention from the public or disparage a rival's good, service, or brand because they are easy to make and require little manual upkeep. An astroturfing and sock puppetry, in general, is unethical and illegal. If detected, sock puppet marketing can have a negative impact, causing potential customers to lose trust and doubt if the product or service is so lacking in value that it cannot be effectively promoted honestly (Wani, M.A., Jabin, S., Yazdani, G. and Ahmadd, N., 2018).

3.2.4 Sybil Accounts

When it comes to sybil accounts, the bad actors make several of them and manage them by hand in order to compromise the reliable network. When a node in online social network claims multiple roles and threatens the security, it is referred to as a Sybil attack. In a Sybil attack, the attacker uses a large number of manually created and maintained pseudonymous identities to spread malware and spam on social networks and gain disproportionately large influence, thereby weakening the reputation of the network. Sybil attackers have many goals like bad mouthing an opinion, illegal voting, accessing resources, compromising security and privacy, etc. Acceding to, social networks with well-defined community structure are more exposed to these Sybil attacks because their links can be used by the Sybil attackers more effectively. Several studies have been carried out so far for the defence of these attacks, but still, the detection of Sybil attacks is in its early stage (Wani, M.A., Jabin, S., et al., 2018).

3.2.5 Bots as Fake Profiles

A bot is a computer programme that uses various scripts to simulate human behaviour online. According to the authors in a bot is a computer program that produces some data to interact with humans especially the persons using the internet (netizens) in order to alter their behaviour? The main use of bots is crawling the data from the web where a simple online computer program identifies and extracts the information from web servers at a much higher speed which was not possible by a human alone. More than 60% of the total web data is generated by bots. But nowadays the bots have been exploited by spammers on different social networks to execute various malicious activities and turned out to be a serious threat to the internet. According to the report, more than 8% bots exist in the Twitter network, and most of them have been developed for commercial purposes. The cyber criminals establish fake profiles on OSNS and control them by an automated program for performing several malicious activities. To spread viral content, bot profiles are used to retweet posts without confirming their origin. Bots are used in online multiplayer games to give players an unfair advantage. Even harder to spot, bots can occasionally create social networks by posing as automated avatars and interacting with people. From the working point of view, bots are similar to Sybil accounts, but the main

difference is Sybil accounts are handled by users manually whereas bots are automated computer programs. To lessen the negative effects of bots, numerous researchers are focusing on their detection. The authors in have deeply studied the behaviour of bot-controlled Twitter accounts and highlighted how bots use different retweet and mentioned strategies while interacting with humans or other bots on the network and presented a framework to detect such accounts. Various OSN service providers employed a number of ways to fight the spam bots. Facebook has its Facebook Immune System (FIS) to deal with bots. However, the users in various OSNs claim that their legitimate accounts are being caught by the detection techniques (Wani, M.A., Jabin, S., Yazdani, G. and Ahmadd, N., 2018).

It is not true that a bot is always designed for malicious activities. They can be used to assist internet users as well. For example, chat bots can be used to help students to answer their everyday queries. Bots which are developed for daily activities like weather update (e.g., Twitter bots) are examples of good bots. But unfortunately, cybercriminals exploit the functionalities of these bots to use them as fake profiles in order to perform various unlawful, misleading, malicious operations. Based on the functionality i present five categories of fake bot profiles as each category is discussed in the following subsections.

Spam Bots

A spam bot is a computer programme created specifically to propagate malicious content, such as links to personal blogs, advertisements, paid content, and pornographic websites. It can also be used to promote any individual or organisation by forging a large number of unwanted connections on the network. In, the authors studied the behaviour of spam bots in the Twitter network and applied several classification techniques to differentiate them from normal bots. One more study in has presented a method known as Bot or Not to differentiate between a human and a bot controlled Twitter account based on six categories of features, viz., network, user, friends, temporal, content and sentiment features.

Although the spam bots are new to the OSNs, the detection of spams has been previously focused on the emails, websites, etc. Spam-bots were first created with the intention of collecting email addresses from the Internet and using them to send unsolicited emails, or spam. According

to the CAN-SPAM Act of 2003, the Federal Trade Commission (FTC) has the authority to levy fines up to $11,000 against business owners engaging in commercial emails. Also, according to Information Technology Act, sending of spam dishonestly or fraudulently is punishable with "imprisonment up to three years or fine up to five lakh rupees or both" (Wani, M.A., Jabin, S., Yazdani, G. and Ahmadd, N., 2018).

Social Bots

Social bots are the computer programs used by humans for their several online activities. According to a study, social bots are highly complex computer programs which behave like humans and usually keep users busy. Social bots are the programs which publicize themselves like viruses to reach and infect a maximum number of users. One more study refers to them as bots which control accounts on online social networks and imitate the behaviour of legitimate users.

Social bots are not always problematic. They are same as other bots in their working, but their focus is more on building social relations with the online people, e.g., Social bots can be used by politicians to engage with the public, by businesses as customer service representatives, by an individual to greatly influence a user or group, and so on. Social bots imitate the human behaviour to gain the attention from their targets (for example, followers, friend requests, replies, likes, etc.) and use this trust network to spread content or promote an agenda or a product. Also, social bots play an important role in multiplayer online games to make the game more entertaining and interesting for the game lovers. Bots can also be used to gain unfair advantages in online games. Establishing and creating social bots is not illegal until and unless it causes any disturbance to the normal functionalities of the system. For example, the IT Act stipulates that the owner of a social bot that spreads malicious content over the network may be subject to "imprisonment up to three years or fine up to five lakh Indian rupees or both."

Like Bots

A "Like" is a support of a post, product, business, etc. registered by clicking the button associated with that item. Like-bots are just computer programs controlled mostly by advertising companies, politicians or normal users to like their products or activities, promote some agenda, etc. One of the main jobs of like-bots is used to increase the 'likes' on ads or pages, but sometimes they can be used to send messages as well. In several OSNs, one can buy fake likes for their content from different online vendors, and the sellers (usually cyber professionals) make use of multiple numbers of like-bots to like the customer content. The number of likes for a product or a page signifies its success and reputation. People use like-bots for their benefit which can misdirect the normal users. Acquiring phoney likes on a product can influence a buyer's perception of its quality, and carefully adding fictitious followers to your page can increase your influence. These days, buying and selling phoney likes and followers can bring in millions of dollars. Online sites like socialbuzzstore15 provide 1000 Facebook page likes for (USD), and five hundred Facebook followers cost 19 USD (Wani, M.A., Jabin, S., Yazdani, G. and Ahmadd, N., 2018).

Thus far, very little research has been done to identify like bots, also known as fake likes. In the authors have conducted a comparative study of 'likes' of Facebook pages produced by Facebook ads and several like-farms. The authors created more than a dozen honey pot pages on Facebook and analysed the produced likes based on users' demography, temporal and social behaviours, etc. Like-farms can make the use of like-bots for their businesses, but naive users need to be aware of these fake likes, otherwise, they are likely to get unacceptable results. One more study analysed a number of Facebook accounts used by some Like-farms and compared their contents (posted on their timeline) with normal user content and found that Facebook accounts owned by like-farms mostly produce likes and comments and often post the same content. Creating like-bots for gaining fake likes and bogus followers deceive the customers and can cause potential customers to lose trust in the organization. These days, it can be very difficult to determine how many "likes" on a post or product are actually from actual users and how many are fake, which can mislead and have an adverse effect on average users. To identify the like-bots and turn them off, a stringent mechanism should be used. The Federal Trade Commission

(FTC) is legally able to impose fines on any corporation proven to be involved in online deceit, including tricking individuals through the use of phoney likes.

Influential Bots

Influential bots are automated identities that illegitimately perform discussions on some trending topics on OSNs like Facebook and Twitter in order to influence opinion or to popularize the topic. Influential bots usually generate messages (tweets or posts) either by reposting (or retweeting) the content posted by other users on the same network or create their synthetic message by an already defined set of rules. The nodes (users) that have the greatest number of connections within the network are referred to as core nodes, and these nodes have a significant impact on a subject or a person. Since one of the goals of influential nodes is to spread the content to the maximum number of people, therefore, they try to send a maximum number of friend/connection requests before spreading the content. The popularity and degree of trust that a node (user) enjoys within the network determines how influential they are; the quantity of incoming requests or received messages determines a node's popularity (Wani, M.A., Jabin, S., Yazdani, G. and Ahmadd, N., 2018).

Influential nodes play an important role in viral marketing, but for marketing companies to identify influential nodes on a network often seems to be challenging. Therefore, nowadays the organizations first design their OSN bots and start getting into online communities to reach the maximum number of customers. These bots begin endorsing goods or brands as soon as they gain the confidence of actual users on the network.

Influential nodes in the network primarily work to sway users' perceptions of a given subject or item. Influential bots, in the same way, try to change the way the people think about an article or any brand on an OSN. Since the normal (real) influential users and the influential bots have almost the same job, therefore it is possible that they have some set of features in common. Using technologies and tools like Klout16 and Twitalyzer17, which are used for regular influential identification, is one potential method of identifying the influential bots. Numerous research projects have been undertaken to determine which online social media users are influential (Wani, M.A., Jabin, S., Yazdani, G. and Ahmadd, N., 2018).

Botnet

The network of automated computer programs in an OSN is referred to as Botnet. In this network, every programme, or bot, is given a different or similar set of tasks to complete automatically. A botnet is a collection of computer programs handled by a 'control-channel' which gives commands to perform unlawful activities. Because the botnet is made up of several bots, the botnet controller can carry out a variety of tasks at once, such as spreading malicious content (spam bot), liking a product or post (like-bot), adding friends to the botnet (interaction/social bot), and gaining popularity for a topic (influence bot).

Botnets are mostly controlled by malevolent users called 'postmasters' by issuing commands to perform malicious activities. In essence, botnets were used to help users in Internet Relay Chat (IRC) chat rooms by moderating conversations, assisting moderators, supplying games, and gathering data about the platform (operating system), and other details of the user such as email addresses, logins, aliases etc. In the authors have studied the growth of social botnet in the Twitter network and observed how the tweets of a normal user differ from the content generated by a social botnet and the ways in which it promotes itself.

Botnets are mostly designed for different kinds of benefits varying from individual to individual, e.g. shopping companies design them to get likes and increase the ratings of their products, researchers, academicians, and data scientists use botnets to crawl data from the web, hackers and other cybercriminals use them as tools for social engineering. A study designed the botnet with three components namely social bots, boaster, and control-and-command-channel which handles the targeted OSN profiles, providing commands (like posting a message, sending friend/connection request, etc.) and carrying the commands respectively. The botnet has been designed to extract the data from the internet. A pictorial representation of a botnet there are three types of users in the diagram viz. normal users, infected users, and bots. Boaster is simply the user (adversary) who owns and controls the botnet and provides the commands via command and control channel. Each bot in the botnet follows the commands. Bots exploit OSNs as an attractive medium to spread the abusive content, bias public opinions, influence user perception and perform fraudulent activities, etc. and are very complex and highly evolving threats to users' trust and security on the internet. As a result, significant methods and actions ought to be done to lessen their impact and the risk involved (Wani, M.A., Jabin, S., et al., 2018).

OSNs are for real people, so managing profiles with automated software is against the guidelines. Bots must abide by the laws and regulations governing the internet, whether they are being used for business, pleasure, or research. Summarizes different kinds of bots and the group of people who mainly use them along with the type of network where they are mostly found. Fake profiles have been seen very risky for both the OSN service providers as well as their users and can be more dangerous in the future if not detected at an early stage. As soon as one creates an OSN account, he/she becomes susceptible to targets of an adversary. The fake profiles can catch one's behaviour and convince the user to perform unlawful activities. From the above discussion, it can be concluded that fake profiles are basically of two types; one created manually and the others using automated methods. Automatic fake profiles pose more threats than other kinds of fake profiles. A boaster can handle several fake profiles simultaneously (botnet) which damages the reputation of the network to a great extent. Therefore, in order to assure the privacy and security of user data and the reputation of the network, it is suggested to focus on characterization and identification of automated fake profiles.

So far, I have seen five different types of fake profiles and their different characteristics exploited by adversaries in order to perform illegitimate activities. Table 1. summarizes these malignant profiles and their main goals on the network along with the group of people/organization that are affected by them.

A study presented a review of the different security and privacy risks to OSN users and endowed with a simple set of recommendations to safeguard user virtual as well as real world.

(Wani, M.A., Jabin, S., Yazdani, G. and Ahmadd, N., 2018) Although, fake profiles are not always harmful; users sometimes create additional profiles for fun and entertainment, or for connecting with specific friend group only, etc. But as they violate the rules and regulations of the service, they are considered illegal. Here rules and regulations in the context of OSN may mean the owner should not have more than one personal account; it should not spread any unlawful or malicious content, it should not collect the user's information or access network by automated means such as bots and spiders18,19, etc. Millions of fake profiles exist on popular social networks like Facebook and Twitter especially in the markets of China and India 20. Social networking service providers are employing a number of ways to ensure user security. For example, Facebook has provided several options to enhance the privacy of user accounts

like protecting the password and sending location-specific login alerts and location alerts. Users can also use the extra security features of the network like how to log- out from remote devices, how to keep Facebook password safe using app passwords, etc. Facebook also has its inbuilt immune system to detect objectionable profiles on the network. Similarly, Twitter and LinkedIn also allow their users to report recognized spam or fugitive content. Recently Facebook introduced an Artificial Intelligence-based system called Deep Text Tool which can understand the text like humans.

	Compromised Profile	Cloned Profiles	Sock Purpose	Sybil Accounts	Fake Bot Profiles
Definition	Existing legitimate profile is taken over by an adversary.	Duplicate profile of existing, legitimate profile created by cloner.	Fake account developed with an intention to deceive others	Multiple forged accounts manually established and controlled by a malicious user	Software program designed to control the fake profile to perform malicious activities automatically.
Purpose	To defame or steal personal information of a person. To spread malicious content by exploiting the trusted network.	To defame or steal personal information of a person. Fun and Entertainment	To honour, defend or support a person or an organization Manipulate a Public opinion.	Bad mouthing an Opinion Casting fake votes To spread malicious content.	To perform viral marketing (influential bots). To increase the number of fake likes.
Types	Partial Compromised (P C), Complete Compromised (C C)	Intra-site cloning, Inter-site cloning	Straw man sock puppet, Meat puppet	---	Spam-bots, Social-bots, Like-bots, and Influential bots

Table 1 Summarization of various Fake OSN Accounts (own source)

Besides helping the users with what they want to say, the tool would also be able to help in filtering the spam content in the near future. Furthermore, Instagram is developing an anti-harassment tool to filter comments or even help users to turn off the comment option on a

particular post. This way the users can block a particular comment to avoid harassment. A study proposed a framework called “Safe book to protect user’s personal data from both the malicious users as well as service providers who violate privacy rules (Wani, M.A., Jabin, S., et al., 2018).

Content-Based characteristics

Content refers to what a user puts or shares on his or her profile, such as text, images, videos, etc. The material reveals a great deal about the user's activity. Not only a user's activity, but also his or her personality as a whole, is represented in his or her network content. Researchers have utilized multiple content-based criteria to identify various types of spammers on various social networks. For the detection of sock puppet accounts on Wikipedia, the authors of employed content-based characteristics such as capitalization, quotation-count, and punctuation marks. Another study on the Chinese social network Sina Weibo used variables such as the quantity of hash tags and URLs in the post to identify network spammers. Similarly, content-based indicators such as the number of images a person has been tagged in and the number of tags in photos uploaded by the user have been used to detect phony profiles on the Face book social network.

Profile-Based capabilities

Profile features refer to the basic information about a user's identification on an OSN, including gender, location, age, phone number, email address, country, profile name, profile photo, number of friends, employment and education, etc. Multiple studies have utilized user profile features to differentiate between normal and abnormal members on various social networks. To identify spammers on the Myspace and Twitter networks, the authors of employed multiple machine learning approaches to profile-based attributes and friend information (such as number of friends, number of follows, etc.) as predictor variables. In, a strategy based on machine learning was used to detect social bots on the Twitter network using profile features such as the ratio of followers to followers of a user's profile. In addition, per Twitter's spam policy, if the number of people following you is less than the number of people you follow, or if you attempt to follow more people than allowed, your identity may be deemed questionable.

3.3 Data collection approaches

The main test in surveying on the web social systems administration locales is accepted to acquire the required dataset. One that is intended for fake profiles, for instance. Researchers have utilized different strategies to accumulate data from ONS locales. An article (Westreich, D. et al., 2010) meticulously describes web information extraction methods and the application spaces in which they are applied.

In this part, I examine many methodologies of extricating the important profile information from social networks. The most well-known strategies for getting the necessary information are building free crawler programs, utilizing as of now accessible advancements to make counterfeit information, and separating information using APIs (Application Programming Points of interaction) given by specialist co-ops.

Data Collection using APIs

The essential use of APIs for information gathering these days is social organization analysis, which is emphatically empowered. To help designers and general clients, OSN specialist organizations much of the time give a scope of libraries (bundles) for different information extraction methods. To get the issue explicit information from a social organization, most scientists compose their own code to interface with it utilizing a Programming interface.

Bot-Based (Crawler) Approach

A part of the bot-based approach is making an autonomous information crawler to recover the information from the social organization. It assembles client information in a way much the same as Programming interface based strategies, but rather than using any APIs, the crawler programs talks with the social organization openly.

Various scripts, like JavaScript, Python, PHP, and so on, can be utilized to foster the information extraction programming. An assortment of seed profiles is vital for any extraction program, however, and these are in many cases chosen in light of various measures, including having an

area based profile and countless companions. For network navigation and information extraction, the application utilizes the seed profiles.

Creation of Fake Data

Systems for get-together data that rely upon bots and APIs consume a huge piece of the day and are very dependent upon client tendencies for security and privacy. A significant part of the time, I truly need data expeditiously to deal with a particular issue, yet this isn't constantly feasible. Furthermore, one most likely will not have the choice to get the important data because of security concerns (Breiman, L., 2001).

In this way, in these cases, I either plan designed data generators or use past data generator gadgets to make the made data test considering the association geology or the features of existing datasets. A number of presently open headways can be used to create the data taking into account the spread out characteristics or estimations of any friendly association that is correct now in presence. For example, if one knows the degree scattering, bundling coefficient, typical between's centrality, and other verifiable estimations, one can construct a phony educational record for investigation. An extent of online data generators can be used to make produced data.

3.4 Methods for profile selection

That's what the previously mentioned subsections clarify, as opposed to the next two methodologies (the crawler-based approach and the Programming interface based approach), current datasets and fake information age techniques require no sort of profile to remove information. Utilizing a Programming interface or a crawler, I want both genuine and counterfeit profiles to extricate information. It is not difficult to distinguish substantial profiles on the social organization since there are such large numbers of them. One can utilize the records of legitimate individuals and their networks of companions to gather genuine profiles. Dependable clients on the organization can incorporate companions, confirmed records, and personal profiles of people you know (Freund, Y. and Schapire, R.E., 1997).

Manual Technique

If I somehow happened to utilize a manual methodology, I would need to investigate any dubious records the hard way and screen any profiles that were viewed as engaged with noxious exercises. While utilizing the manual choice technique, there are multiple ways of looking at and pick the bogus profile set. A manual rundown of randomly chosen profiles from an organization can be made, and each profile can then be named in view of a bunch of qualities that recognize certifiable profiles from bogus ones. For instance, data might be manufactured and later got from a client's or alternately companion's profile through one of the information assortment draws near assuming it is observed that the individual is sharing destructive or unlawful substance on their profile.

Honey Profile-Based Approach

Honey profiles are OSN profiles that are intended to draw in different clients, undoubtedly those with comparable interests. Different sorts of "honey profiles" or "honeytraps" are created when important to draw in both fair and deceptive individuals. For instance, some make honey profiles to interest a particular segment, similar to teenagers and youthful grown-ups on the organization they are focusing, while others do it to speak to a more extensive crowd. Nonetheless, with

regards to choosing fake records, the analysts deliberately make honey profiles — like those that are express — to attract the fake records that fall into a similar class.

To keep the honey profiles new and connecting with, the proprietors consistently add new and enthralling narratives and images. The writers made more than 900 honey profiles on Twitter, Facebook, and Myspace. In the wake of social affair these profiles, they glanced through them to search for any odd action from clients who drew in with them (Williams, C.K. and Barber, D., 1998).

Botnet-Based Approach

A botnet, as was recently referenced, is a gathering of robotized projects, or bots, that are constrained by a solitary human administrator known as the "botherder." The bots are modified to play out a scope of capabilities, for example, cooperating and attracting different clients, publicizing items and administrations, directing political races, etc. Rather than choosing prior counterfeit profiles on the organization as I did in the past two ways, the botnet-based methodology includes infusing a few computerized, intuitive misleading profiles into the organization and expanding client trust levels to secure the rundown of phony profiles.

Dissimilar to honey profiles, which for the most part stand by latently for the predetermined association demands, profiles in a botnet are normally engaged with exercises, for example, giving association demands, conveying material, and so on to grow an enormous client base. A fake explicit profile or another profile that disseminates grown-up happy is one sort of profile that the honey profile-based strategy is fruitful in attracting, yet it is less effective in drawing in astroturfing accounts.

3.5 Social media network

History and Evolution of social media platforms

The last decade has shown a continuously increasing graph in the usage of Social media networks especially LinkedIn, Face book and Twitter. As of 2018, Face book has more than two billion users and Twitter has more than 262.2 million users. Social media networks are expanding largely due to high internet availability and increasing Literacy rates. Through these social networks, communicating around any edge of the globe has been easier than ever. Social media networks have not only improved communication skills, it has promoted education, business, employment, skills, etc. on a larger scale.

With the increasing evolution of social networks, many intense serious issues have raised like online impersonation, fake accounts, cybercrimes, sexual harassment, cyber bullying, privacy and security concerns, etc. A fake account is the account which is not owned by any real person or organization, rather it's been created for getting malicious activities done. Apart from fake accounts, clones also exist which are real people doing fraud or spam activities. A lot of information is available on a person's account which includes details like photographs, name, interests, attire, education, e-mail ids, contact, etc. By this information, a person's general view gets imaged in the subconscious and it becomes easier to make a fake account taking data of any individual (Smola, A.J. and Schölkopf, B., 2004).

These kinds of activities or say attacks need detection techniques for being aware of such cases at an early stage and protecting ourselves from all kinds of spams and scams. Detection techniques play a major role in finding such fake profiles which gather people's information by gaining trust or by using wrong measures. Fake profiles and clones have become a major threat since the last few years. Many authors and researchers have worked on it and suggested ways like detection through machine learning, examining algorithms, etc. to avoid such attacks.

Crucial issues arising due to social networks

(Jeatrakul, P. and Wong, K.W., 2009) Detection of such fake profiles is a very challenging task among the huge source of web, data and links available. The major issues that are arising due to social media networks can be examined in following points:

1. People should not reveal their true identity on social media platforms according to privacy policies. The authenticity of true identity has been questioned very often and this affects the people who are falsely accused or been misled by resources.
2. The ease with which fake accounts and malicious activities are happening around is among the major issues. False accounts being bought at online marketplace at minimal costs. It has been so easy to buy fake followers and like for social sites like Instagram, Face book and Twitter, etc.
3. The miss-authentication of social media sites by having more followers, likes and comments leads to gaining more attention and social popularity. This trend motivates individuals to seek out new artificial or manual means to keep ahead of the competition.
4. It is well-known that social media platforms are susceptible to a new term known as a Sybil attack, in which an attacker or fraud maintains a large number of phony accounts and utilizes them to conduct a variety of hostile acts.
5. Groups which are especially made to spread chaos and miss-conception on social issues or for spamming activities. Fake news circulated across social media about Hurricane Sandy in the US. The news became viral in such a short time in spite of being fake and became a major source of information for those who got affected by the storm.

According to current scenario and graph analysis, most of the audience uses manual methods to detect fake accounts. A fake account doesn't use genuine words; it gets more involved into persuasive words. Fake accounts can be identified by analysing the algorithm and actions followed by any account. Bogus up-sides can possibly truly disable clients' encounters, consequently quick move ought to be made to suspend accounts that are believed to be deceitful. Social media networks haven't totally mechanized the most common way of removing false records.

3.6 Detection of fake account using machine learning

Nowadays, social media has become the most facile platform for doing fraud or spreading distorted news. Threats like spamming; identity impersonation and social abuse are issues attaining high rates with time. None of the social media platforms is yet so secure that you can trust on private issues. Each data is stored and reviewed by some people around firms which manage the concerned department of that online social network. Analysing the data stored, a lot of information is gained to understand what content people like to have and how people do react on different issues. These measures are used to provide better services, for improvement in their network to get more engagement by the audience. But the same applies to the other side, a lot of people can access the data provided on these platforms through dark measures. And thereafter the data can be used for blackmailing, fraud, abusing on certain terms, spamming, and many such problems (Toloşi, L. and Lengauer, T., 2011).

To detect such issues, Researchers have worked upon it and Machine learning methods have been introduced. Machine learning algorithm should be well known to understand the user profile. Machine learning is used to detect BOTs on social media and fake account activities are also detected through it. It also detects the purpose behind any bot. I can recognize fake pages and records by utilizing a system that utilizes a component to distinguish threatening bots, which eventually transform into false records. The central concern with these bots going about as fake social media accounts is that they can spread all over by posting bogus data and creating characters to deceive individuals about connections, finances, and different parts of their lives. To remove the required data, this approach centers principally around Twitter and its URL attributes. The review was directed with thought for the clients' believability and the protection of their tweets, guaranteeing that each record is considered reliable freely. Following the extraction of the fundamental information from the Twitter URLs, a Learning Automata calculation is utilized to decide if a record is that of a phony client or a bot on the social media organization. AI is utilized in this manner to distinguish perilous bots on social media sites and accumulate information from them.

3.6.1 Understanding Theory and Intuition behind Neural Network

Neural Mathematical Model

Data is accumulated by the neuron from input channels called dendrites, handled in the core, and result created in a long, slender branch known as an axon. Consists the three main aspects, that are Dendrites, Nucleus and Axon these are considered as independent variables, Processing operations and Activation function respectively. Considering the independent variables as X2, X2, X3. Xn Operations processor as some 'P', and activation functions as 'F'. Some of the Activation Functions are:

Sigmoid: acknowledges a worth somewhere in the range of 0 and 1. It changes enormous positive numbers to one and large regrettable numbers to nothing. Regularly, it fills in as the result layer. Notwithstanding, SoftMax is what I use as the result layer in our model.

SoftMax: The multi-class calculated relapse or delicate argmax capability are different names for the SoftMax capability. This is because of the way that the SoftMax's recipe is very like the sigmoid capability utilized in strategic relapse, making it a speculation of calculated relapse reasonable for multi-class characterization.

Since it changes the scores into a standardized likelihood conveyance that can be displayed to clients or took care of into different systems, the SoftMax is unbelievably useful. Along these lines, the SoftMax capability is ordinarily added as the brain organization's last layer.

RELU- Rectified Linear Units: A non-straight enactment capability called ReLu is utilized in profound brain networks and multi-facet brain networks. Alternatively, in profound brain organization or traditional brain network ideal models, the actuation values have been registered utilizing the ReLu capability. Since the subordinate of the ReLu capability is 1 for a positive info, it can accelerate the preparation of profound brain networks more rapidly than traditional enactment capabilities. commonly utilized in secret layers.

EXAMPLE

Allow us to consider an illustration of a unit step enactment capability. The information is planned somewhere in the range of 0 and 1 utilizing these actuation capabilities (0,1).

Ann Training Process

Seldom is ANN used in prescient displaying. Artificial Neural Networks (ANN) regularly endeavor to over-fit the relationship, which is the explanation. ANN is ordinarily used in circumstances where authentic occasions are almost indistinguishable from each other.

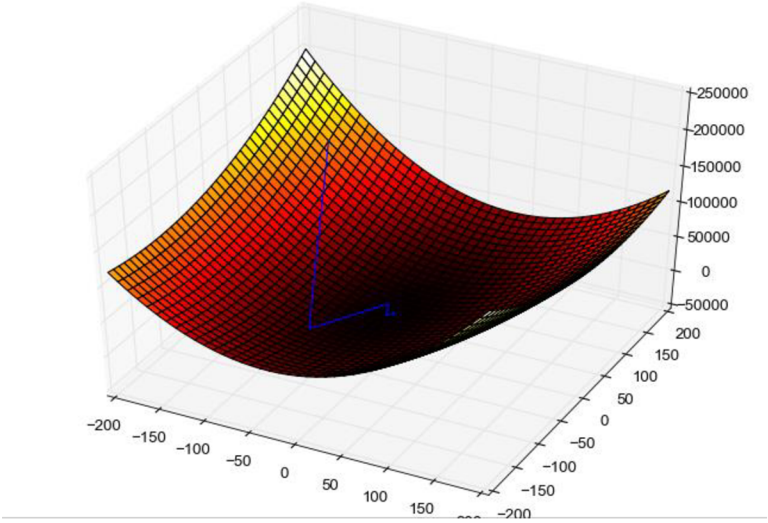

Figure 1 Gradient of ANN process (own source)

To find the ideal organization weight and predisposition values, one enhancement strategy is slope plunge. It endeavors to iteratively limit the expense capability. It figures the expense capability's slope and moves in a negative heading until the neighborhood or objective least is reached. At the point when the inclination's positive worth is utilized, a neighborhood or worldwide most extreme is reached.

Back Propagation

ANNs are prepared utilizing back proliferation, which processes the slope expected to refresh network weights. It is habitually utilized as an advancement approach for inclination plunge to change the heaviness of neurons by sorting out the angle of the misfortune capability.

PHASE 1:

proliferation forward to create the result value(s) over the organization. Cost calculation (blunder term). Deltas are delivered by spreading yield enactments back through the organization utilizing a preparation design focus to make variations between the expected and real result values).

PHASE 2:

Weight revisionWeight slope estimation.

This proportion, frequently known as the learning rate, influences the rate and type of learning. Neurons train all the more rapidly at higher proportions, however more precisely at lower proportions.

4. Practical Part

The methodology for machine learning in detecting fake accounts on social media platforms like Instagram and Twitter involves a structured approach comprising four primary stages. Firstly, in the profile categorization stage, the dataset is systematically organized based on various features extracted from the profiles. These features include metadata and multimedia content, and techniques such as similarity assessment and facial recognition analysis are utilized to differentiate between genuine and fake accounts. Following this, in the data collection phase, extensive efforts are made to gather relevant data from various sources, employing methods like web scraping and sorting algorithms.

Profile Categorization

In light of the sight and sound substance and metadata of the distributed work, highlights have been recovered from the picked dataset. A couple of AI parts that go past extraction are: Using similitude assessment procedures between two-character strings or words, nickname contains name (ncm) is executed; in this case, the usernames and names are considered.

The accompanying segments are shown along with their rate proportions; the element percentage_completed_profile is gotten from the extensive analysis of these few segments;

1. Has at least one post (10 %)
2. Has 10 or more posts (10 %)
3. Has at least one follower (5 %)
4. Follows at least one account (5 %)
5. Has name (5 %)
6. Has a description (15 %)
7. Has a web page (5 %)
8. Has a profile picture (20 %)

9. Have face in the profile photo (10 %)
10. A language was detected (15 %)

Photos similarly this feature takes a record of each picture and analyses, confirming if there are any known faces and marked as similar if known faces are found.

Data Collection

For data extraction on third parties and irrelevant pages linked to Facebook or Instagram, sorting algorithms were employed. For the web scraping and sorting section, Python and Selenium were the predominant technologies employed. For additional publishing analysis, Google Vision APIs are utilized. Now, an original page, fan pages, and phony ones exist. Followers of official pages of the programming league are taken into a role for analysing real profiles and then, if any profile is found, it gets verified manually. Fake profiles and impersonating pages are published as fake after extraction from verified forums. Each technique result is verified manually at the end.

Data Scrapping

I initiated the process by creating a Twitter Developer Account through the Twitter Developer Portal, where I acquire essential credentials such as the API key, API secret key, Access token, and Access token secret. Once these credentials are obtained, I proceed to set up API authentication using a library like Tweepy in Python, which efficiently handles OAuth 1.0a authentication. I integrate these credentials into my code to authenticate my application. With authentication established, I am ready to commence making API requests utilizing the authenticated API object. I specify various parameters, including user profiles, tweets, trends, and filters like tweet count and date range, to precisely retrieve the desired data.

To prevent potential abuse, it is vital for me to effectively handle rate limiting imposed by the Twitter API. This entails implementing appropriate waiting times or employing backoff strategies to ensure compliance. Upon making requests, I proceed to extract and process the data retrieved from the Twitter API's JSON responses. I parse these responses to extract the necessary

information, which can then be stored locally, analyzed, or further processed based on my specific requirements. I am able to access structured data provided by the Twitter API, enabling me to retrieve specific information about users, tweets, trends, and more, in real-time. For seamless integration of Twitter functionality into my applications, I can utilize developer tools like Tweepy, which streamline the process of incorporating Twitter data. Customizing data retrieval is made possible by leveraging parameters and filters to tailor the data retrieval process to suit my specific needs and preferences. To ensure compliance with Twitter's terms of service and mitigate potential legal risks associated with data collection, adherence to platform guidelines is essential.

Feature Selection

Information change is finished all through the element choice cycle by changing nominal information into numeric Boolean sections. Then, a conventional scaling was utilized, which included taking the mean out and scaling to the change of the unit z = (x u)/s, where s is the standard deviation and u is the mean of the preparation test. A couple of qualities are removed for this determination, and as they are nominal and utilized in numerous different examinations, they are not vital for this emphasis. Utilizing the leftover characteristics, Pearson's vicariate connection procedure was utilized to direct a relational analysis that can be seen as a proportion of the connection between the factors.

4.1 Machine Learning algorithms

To assess all the varied outcomes, numerous tests were conducted with the following categorized methods in mind: accuracy, precision, unpredictability, and size. Eighty percent of the dataset was used as training data and twenty percent was randomly selected to test the model.

Random forest

Permutation and combination algorithm introduce randomness and creates a set of classified groups as a key factor. Randomness is introduced here to reduce the estimator's variance. By mixing numerous trees, random forests achieve reduced variation, sometimes at the expense of a modest increase in bias.

Decision Tree

A widely used technique these days builds a model based on a set of rules and forecasts the value of the target variable. Although this method produces a number of complex trees, it functions pretty appropriately. That makes the data arranging complex and this method is termed as over fitting. In the implementation, over fitting problems gets resolved.

Quadratic Discriminate Analysis

It is a traditional system that provides straightforward implementations because it lacks hyper-parameters for adjusting the analysis. Depending on the nature of the problem, a suitable solution may be provided. Due to the non-linearity of certain variables, this approach produced the worst results during implementation.

Gaussian process classification

This method works upon Gaussian probabilistic prediction which comes under unsupervised learning algorithms. These algorithms lose efficiency in some implementations since they are very tactful in big spaces, and the outcomes are not always the best.

4.2 Detection of fake accounts on twitter using ML

With Twitter being perhaps of the most utilized social medium stages out of the multitude of sites, social media has become unbelievably well known. Starting around 2018, there were 262.7 million users. To battle misrepresentation and other undesirable demonstrations, Twitter has executed various techniques and endeavours. Detailing and forbidding somebody who takes part in such way of behaving is a fabulous mediation. In any case, to try not to rely upon the client base stages for a drawn out arrangement, a computerized identification model was genuinely required. Thus, the making of mechanized identification models includes AI. A lot of exploration has been finished, including random strolls and studies, to all the more likely comprehend and foster record identifiers that influence the profiles of AI clients.

To recognize sham and genuine records autonomously, I can use graphical and analytical procedures. Albeit vindictive records will quite often utilize influential and cruel language, genuine records will quite often utilize more formal and normal language.

1. One study focused on Twitter accounts that predominantly retweet Italian politicians and engage in various forms of advertising, such as on Amazon.com or job postings. This analysis relied solely on evaluating existing techniques, achieving a maximum accuracy model through the use of an Activity DNA approach, which involves analyzing sequences of account activities including tweeting, retweeting, mentioning, hash tagging, and posting pictures or URLs. The model achieved an impressive accuracy rate of 98% using a rule-based approach called 'Longest Common Substring'.

2. Another analysis utilized Google Safe Perusing and Catch HPC to recognize Twitter accounts that contained tweets with possibly destructive connections. 500 records were named spam bots and 1500 as genuine records by this review. Various pointers were analyzed, including the accompanying: following speed, normal adjoining tweets, normal adjoining devotees, betweenness centrality, bidirectional proportion, and grouping coefficient. The review utilized a Random Forest calculation to accomplish an exactness pace of 85%.

3. An other report isolated 500 spam bot accounts from 500 genuine records by seeing Twitter accounts that common spam content, like adverts or malevolent connections. This analysis considered factors including the proportion of companions to devotees, the amount of messages conveyed, and the extent of tweets that contained URLs. A 2.5% misleading positive rate was accomplished by the review utilizing the Random Forest strategy.

4. Another investigation targeted Twitter accounts involved in purchasing fake followers. This study categorized 1950 accounts as having fake followers and 1950 as genuine. Factors considered included whether the account had at least 30 followers, whether it was favourited by other accounts, whether it used hashtags at least once, and whether it had posted at least 50 tweets. This study achieved an accuracy rate of 99% using a Random Forest algorithm.

5. In a different domain, a study analyzed reviews on the Google Play Store, distinguishing between spam reviews and genuine reviews. Measurements including feeling extremity, normal letters per word, rating, length of surveys, and cosine closeness of text bodies were considered in this analysis. The review's precision rate was 97%.

6. Records can be separated into two classes: multiclass and binary. The binary class for the most part comprises of spambots and counterfeit adherents, the two of which can take part in serious criminal behaves like misrepresentation and spamming. Three particular types of multiclass work are recognized: devotees, spambots, and genuine records.

Fake and clone profiles are becoming a very major threat with each passing day. Social media gathers a lot of information like image, contact details, email id, school/college name, address, location etc. This information is readily exposed in social platforms for several reasons, hackers can get advantage of this data to create fake or clone profiles. They can use these fake and clone accounts to cause problems like phishing, spamming, fraud activities, spreading fake news, cyber bullying etc. Sometimes, they even try to defame the real owner or the account or organisation.

4.3 Analysis

In this study I will create a model to detect fake Instagram profile and study the efficiency of the presented model

4.3.1 Instagram fake accounts identifier

Comprehending the Issue

The objective of this examination is to make and sharpen a profound brain network model for distinguishing fake or deceitful Instagram accounts. Spam accounts are turning into a major issue on all social media stages nowadays. To give the feeling that they have an enormous number of supporters on their own records, a few group are making fake profiles. To advertise fake labor and products, false records are being made.

Additionally, they are being used to impersonate other record clients, going from customary individuals to celebrities, with an end goal to put individuals in a terrible mood and notorieties and apply impact. I thought about a couple of critical information boundaries to survey whether the record is fake.

The features of the input are

PROFILE PICTURE whether the client has transferred one. The proportion of the quantity of numeric characters in a username to its length is shown as NUMS/LENGTH USERNAME. Complete names are recognized in word tokens by the FULLNAME WORDS feature.NAME AND Complete NAME LENGTH the extent of mathematical characters to length in a complete name. Are the username and complete name the very same, as the NAME capability is intended to recognize USERNAMEs? The bio length in characters is perused by the Depiction LENGTH input capability. Additionally, it checks whether it has an external URL utilizing the EXTERNAL URL functionality. whether the profile is private. It explores the amount of devotees, posts, and follows.

Trained Detector Model

This model is prepared to assess the qualities recorded above and decide if a given record is false. By giving the result a worth of 0, which shows TRUSTED, or 1, which demonstrates Counterfeit. I want to enable this product to have a similar outlook as a human, settling on choices in view of the data gave and boosting the probability of progress.

576 rows 12 columns	575	574	573	572	571	…	4	3	2		0
profile pic	1	1	1	1	1	…	1	1	1	1	1
nums/length username	0.27	0.57	0.57	0.38	0.55	…	0	0	0.1	0	0.27
full name words	1		2			…	2		2	2	0
nums/ length fullname	0	0	0	0.33	0.44	…	0	0	0	0	0
name == username	0	0	0	0	0	…	0	0	0	0	0
description length	0	11	0	21	0	…	0	82	0	44	53
external URL	0	0	0	0	0	…	0	0	0	0	0
private	0	0	0	0	0	…		0	1	0	0
#posts	2	0	4	44	33	…	6	679	13	286	32
#followers	150	57	96	66	166	…	151	414	159	2740	1000
#follows	487	73	339	75	596	…	126	651	98	533	955
fake	1	1	1	1	1	…	0	0	0	0	0

Table 2 Datasets (own source)

Conducting Investigative Data Analysis For The "Training Dataset"

	Count	mean	Std	min	25%	50%	75%	max
profile pic	576.00	0.70	0.46	0.00	0.00	1.00	1.00	1.00
nums/ length user	576.00	0.16	0.21	0.00	0.00	0.00	0.31	0.92
Full name words	576.00	1.46	1.05	0.00	1.00	1.00	2.00	12.00
nums/ length full	576.00	0.04	0.13	0.00	0.00	0.00	0.00	1.00
name == username	576.00	0.03	0.16	0.00	0.00	0.00	0.00	1.00
description length	576.00	22.62	37.70	0.00	0.00	0.00	34.00	150.00
external URL	576.00	0.12	0.32	0.00	0.00	0.00	0.00	1.00
private	576.00	0.38	0.48	0.00	0.00	0.00	1.00	1.00
# posts	576.00	107.47	402.03	0.00	0.00	9.00	81.50	7389.00
# followers	5.76	8.53	9.10	0.00	3.90	1.50	7.16	1.53
#follows	576.00	508.36	917.96	0.00	57.50	229.50	569.50	7500.00
fake	576.00	0.50	0.30	0.00	0.00	0.50	1.00	1.00

Table 3 Training Dataset (own source)

Performing Exploratory Data Analysis

This model is trained to evaluate the aforementioned characteristics and decide the authenticity of an account. The output will be either 0 or 1, denoting TRUSTED or FAKE, accordingly. Our goal is to enable this software to make decisions similar to those made by a human, based on the information provided and leading to the highest possible likelihood of success.

Number of unique values in the profile pic column

Starting with the analysis of profile pictures, it's noted that out of 576 total accounts (404 trusted and 172 fake), there are 404 accounts with profile pictures among the trusted accounts, while only 172 fake accounts have profile pictures. This indicates that a higher proportion of trusted accounts have profile pictures compared to fake accounts. Moving on to the breakdown of fake and trusted accounts, the data reveals an equal distribution of 288 accounts each in the trusted and fake categories. This suggests a balanced dataset with an equal representation of both types of accounts. Considering the presence of external URLs, it's observed that out of 576 total accounts, 509 trusted accounts have external URLs linked, while only 67 fake accounts have external URLs. This indicates that a significantly higher number of trusted accounts include external URLs compared to fake accounts.

Performing Data Visualization

In the presented figure 2, each data point represents a profile, and the graph visualizes whether each profile is categorized as "trusted" (0) or "fake" (1). The horizontal axis likely represents individual profiles, while the vertical axis indicates the classification label assigned to each profile. By analyzing the distribution of data points along the vertical axis, one can determine the proportion of profiles categorized as trusted (0) versus fake (1). This visualization offers a quick and intuitive way to understand the distribution of fake and real profiles within the dataset.

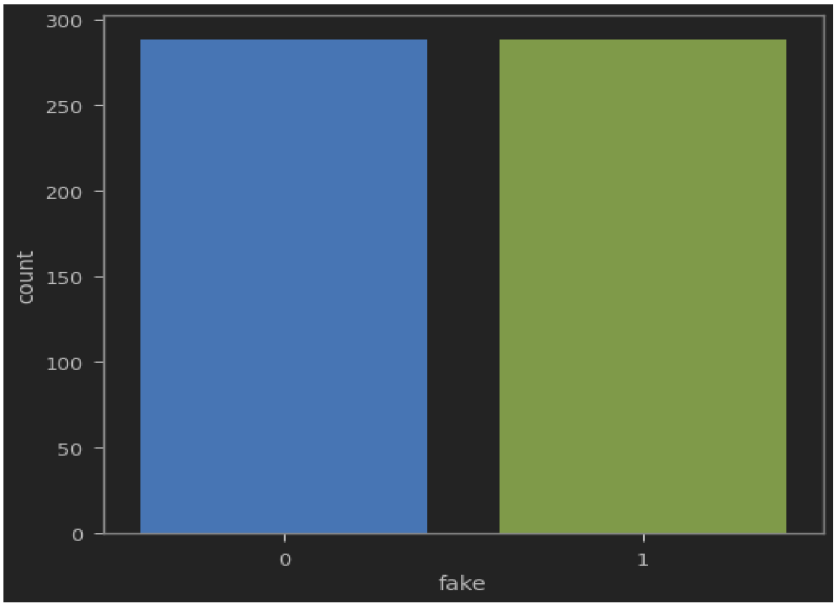

Figure 2 Visualizing the number of fake and real accounts (own source)

The figure 3 here represents the privacy of profile, number of profiles has made their setting on private profiles It can be drawn that over 350 profiles are trusted profiles and 200 are fake who have made their profiles private.

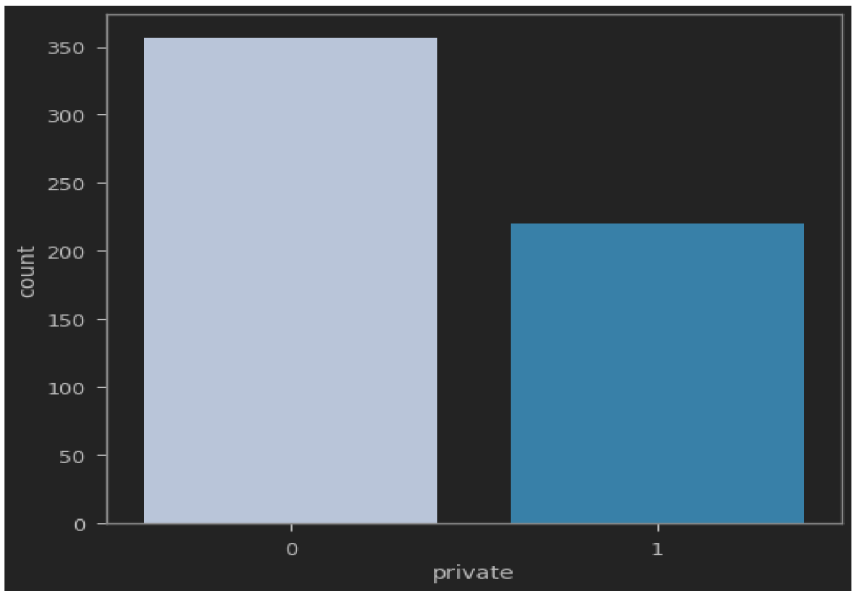

Figure 3 Visualizing the private column (own source)

The figure 4 visualizes that over 400 profiles who have set their profile picture are fake and more than 150 who have profile pic are trusted.

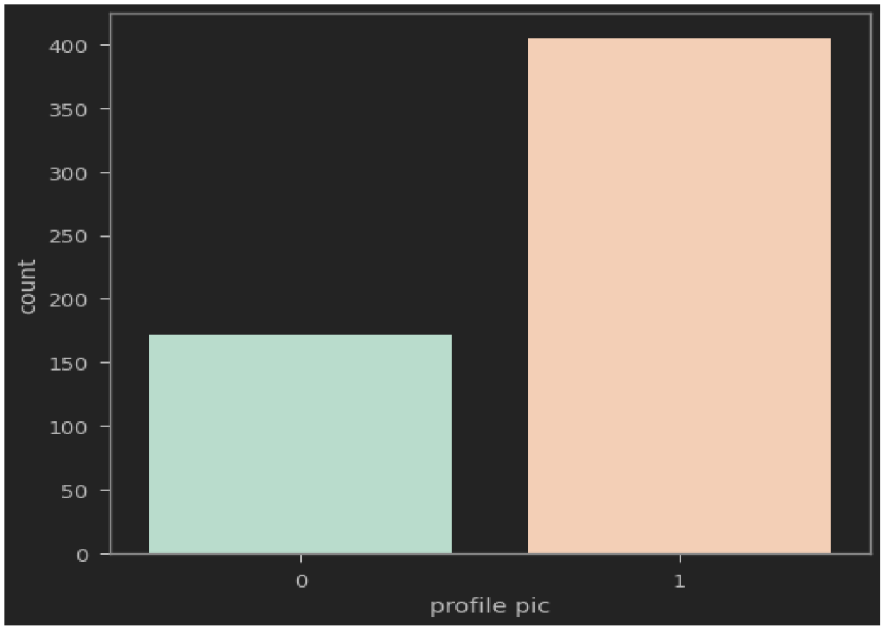

Figure 4 Visualizing the profile pic feature (own source)

The figure 5 here depicts the length of username of the profiles who been set to analysis.

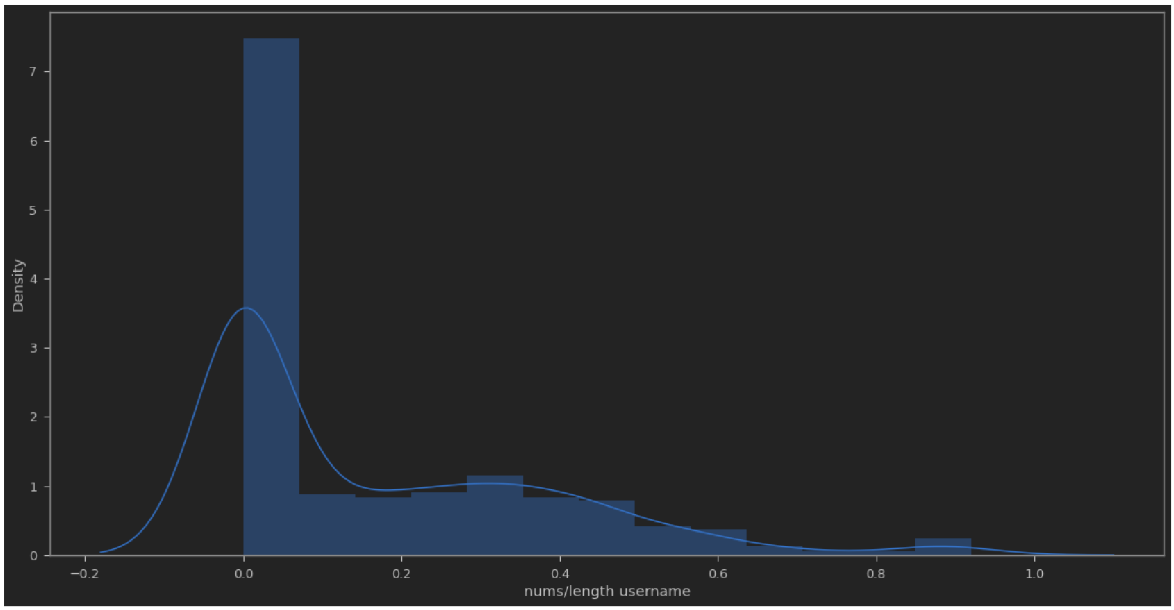

Figure 5 Visualizing the length of usernames (own source)

Correlation heatmap

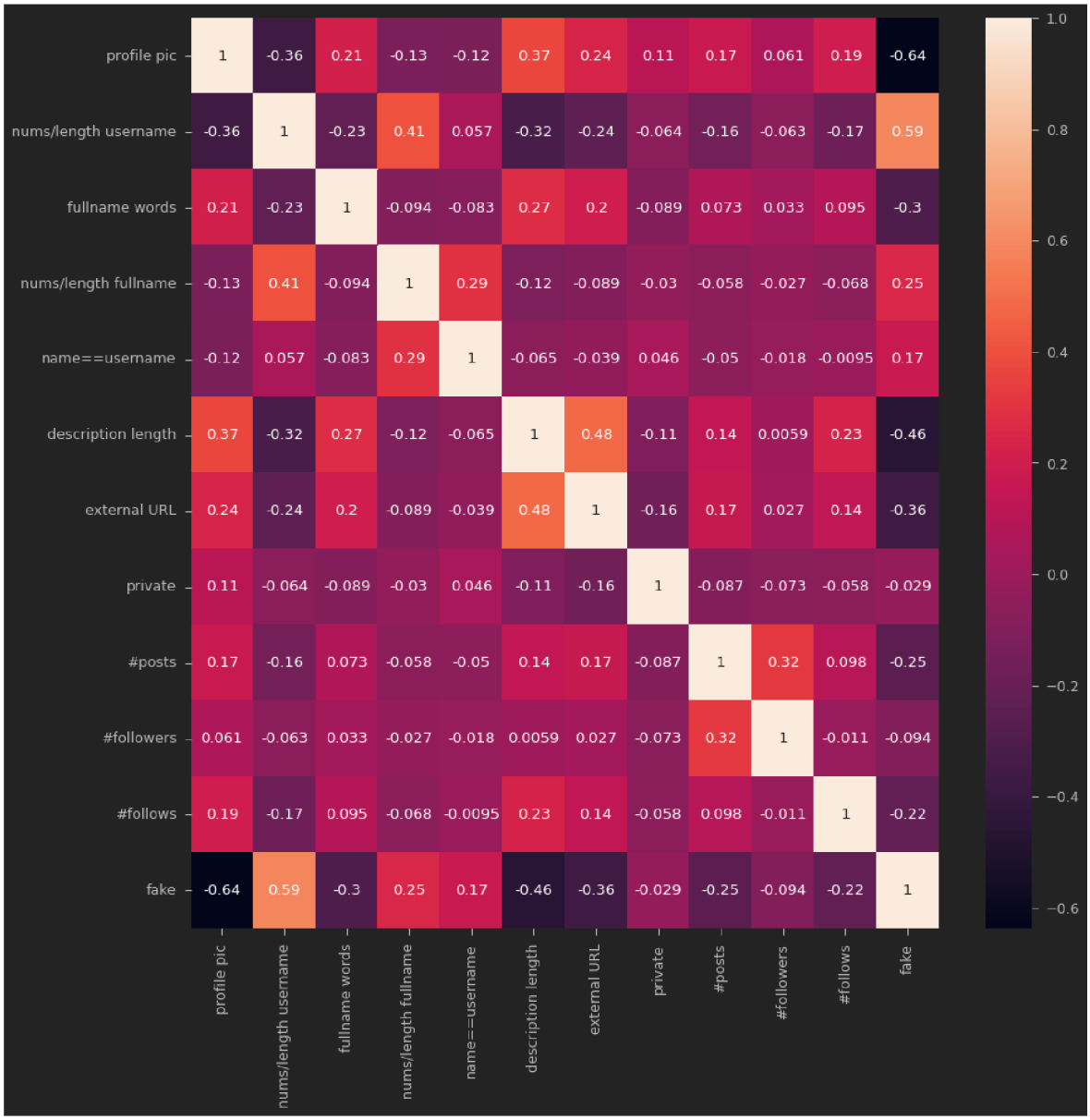

Figure 6 Confusion Matrix (own source)

4.3.2 Implementation and Analysis of Neural Network Models

Build A Simple Deep Learning Model

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 50)	600
dropout (Dropout)	(None, 50)	0
dense_1 (Dense)	(None, 150)	7650
dropout_1 (Dropout)	(None, 150)	0
dense_2 (Dense)	(None, 25)	3775
dropout_2 (Dropout)	(None, 25)	0
dense_3 (Dense)	(None, 2)	52

Total params: 12,077
Trainable params: 12,077
Non-trainable params: 0

Table 4 Deep Learning Model (own source)

Assessing The Performance of The Model

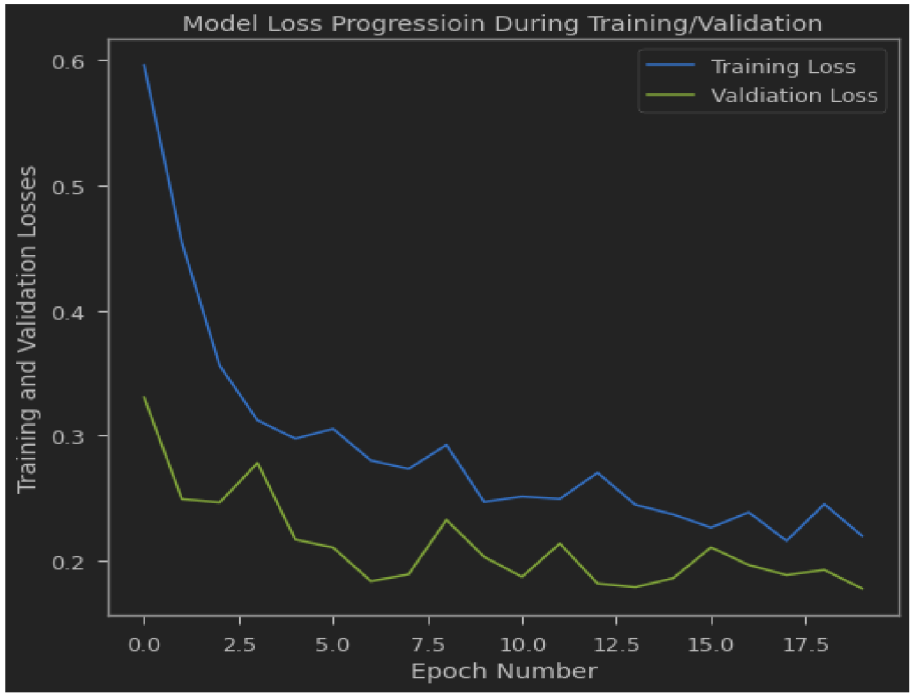

Figure 7 Graph represents the loss of model at the time of training (own source)

Specifically, I need to focus on memory, exactness, and accuracy as well as the F1 score. The proportion of all accurately expected cases — positive or negative — to the all out number of cases in the information is exactness, an extremely fundamental presentation marker.

The proportion of accurately anticipated positive cases to precisely anticipated positive cases is known as accuracy. It answers the accompanying inquiry: Which level of the expected patrons in the data set were genuine donations? High accuracy ought to, in principle, lead to a low bogus positive rate. This is significant on the off chance that our financial plan for our post office based mail crusade is restricted on the grounds that i would rather not send mixed up mail to individuals who will not answer! To upgrade our income, accuracy is essential.

The proportion of accurately recognized positive examples to the complete number of genuinely certain cases is known as review or responsiveness. It answers the accompanying inquiry: How much did our model accurately identify supporters among every single genuine contributor? This is critical in light of the fact that i need to reject whatever number genuine contributors from our mission as could be expected under the circumstances to augment in general reaction (or pay)!

As a weighted normal of review and accuracy, the F1 Score represents both bogus up-sides and misleading negatives. In situations where there is a lopsided appropriation of classes in the information, the F1 score holds impressively more noteworthy worth than precision. Precision is expanded when the expenses of misleading up-sides and bogus negatives are same. It is smarter to consider accuracy and review on the off chance that there is a tremendous distinction in the expense of bogus up-sides and misleading negatives.

	precision	recall	f1-score	support
0	0.86	0.92	0.89	60
1	0.91	0.85	0.88	60
accuracy			0.88	120
macro avg	0.89	0.88	0.88	120
weighted avg	0.89	0.88	0.88	120

Table 5 Accuracy Score (own source)

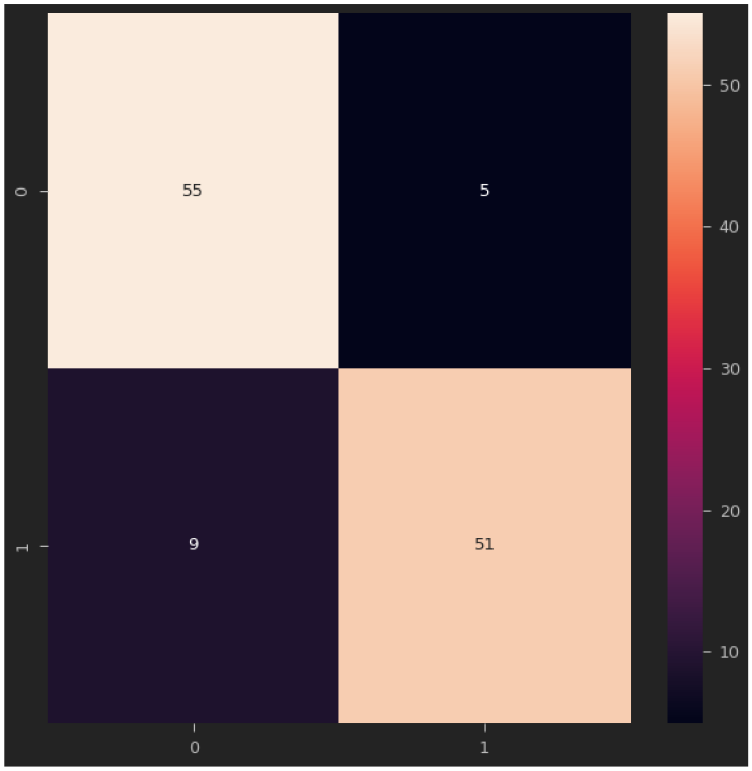

Figure 8 Confusion Matrix (own source)

5. Results

I have analysed the dataset and obtained a fairly accurate predictive model using Neural networks. The model is hence trained to detect fake accounts in Instagram based on the considered features. I achieved 95 percent accuracy in detecting the fake accounts by training the model using datasets (from train.csv). I have checked whether the model has reached the ability to detect an account is fake or not by inputting different set of data values (test.csv file) which consisted of 120 account details. The model predicted true values for 106 accounts and predicted false values for 14 accounts out of 120 accounts. The analysis of the dataset and the subsequent development of a predictive model for detecting fake accounts on Instagram represents a significant advancement in the realm of social media cybersecurity. Leveraging neural networks, I have constructed a robust model trained on a carefully curated dataset to accurately identify fraudulent accounts based on a range of pertinent features. The culmination of this effort has yielded a predictive model with an impressive accuracy rate of 95%.

To achieve this milestone, my methodology involved extensive preprocessing of the dataset to ensure its integrity and relevance for training the model. I then proceeded to train the neural network using the training dataset (train.csv), which comprised a comprehensive set of account details encompassing various attributes and characteristics. Through iterative training iterations, the neural network learned to discern patterns and correlations within the data, ultimately gaining the ability to differentiate between genuine and fake accounts with a high degree of accuracy.

Following the training phase, I conducted rigorous testing of the model's efficacy using an independent dataset (test.csv), which contained 120 account profiles distinct from those used for training. This step was crucial for evaluating the model's generalization capability and its performance on unseen data. Remarkably, the model demonstrated remarkable proficiency, correctly identifying 106 out of 120 accounts as either genuine or fraudulent. The model's predictive accuracy was further validated through the identification of 14 accounts falsely classified as either genuine or fraudulent. This outcome underscores the importance of continued refinement and optimization of the model to minimize false positives and negatives, thereby enhancing its reliability and effectiveness in real-world applications. The success of our neural network-based approach in detecting fake accounts on Instagram holds significant implications

for enhancing cybersecurity measures on social media platforms. By leveraging advanced machine learning techniques, I can proactively identify and mitigate fraudulent activities, thereby safeguarding user trust, privacy, and security in online environments.

The insights gained from this study provide valuable guidance for platform operators, cybersecurity professionals, and policymakers seeking to combat the proliferation of fake accounts and deceptive practices on social media. The development of robust detection mechanisms informed by data-driven approaches can serve as a critical line of defense against malicious actors seeking to exploit vulnerabilities and manipulate online discourse for nefarious purposes. Moving forward, there are several avenues for further research and development to build upon the foundation established in this study. For instance, the incorporation of additional features and data sources, such as user engagement metrics and content analysis, could enhance the model's predictive capabilities and resilience to adversarial attacks. Additionally, exploring ensemble learning techniques and hybrid models may offer synergistic benefits in terms of accuracy and robustness. It is essential to consider the ethical implications and societal impact of deploying automated detection systems for identifying fake accounts. Striking a balance between security imperatives and user privacy rights is paramount to ensuring the responsible and equitable use of technology in combating online threats.

This study aimed to address the escalating concern surrounding the proliferation of spam accounts across various social media platforms, particularly on Instagram. These fake profiles serve various nefarious purposes, including artificially inflating follower counts, promoting counterfeit products, and impersonating legitimate users for malicious intents such as influencing, criticism, and reputation damage. To combat this issue, I devised a robust model leveraging deep neural networks trained on a multitude of critical input features characteristic of both genuine and fake Instagram accounts. These features encompassed a comprehensive array of parameters, including the presence of profile pictures, numerical attributes of usernames, bio length, existence of external URLs, privacy settings, and metrics related to post, follower, and following counts. Using a meticulously curated training dataset, I meticulously fine-tuned the detector model to accurately classify accounts as either trustworthy (output: 0) or fraudulent (output: 1) based on the discerning features identified. My primary objective was to instill within the model the cognitive capacity to emulate human judgment, thereby maximizing its efficacy in discerning between genuine and fake profiles. Through extensive exploratory data

analysis, I meticulously scrutinized various facets of the dataset, ranging from the distribution of profile pictures and presence of external URLs to the privacy settings across trusted and fake accounts. These analyses were bolstered by intuitive visualizations, including insightful graphs, which facilitated the elucidation of prevailing patterns and trends inherent within the dataset. Furthermore, I elucidated the theoretical underpinnings of neural networks, elucidating the intricate mathematical formulations governing neurons, activation functions, and the iterative training process encompassing gradient descent and backpropagation. This theoretical exposition served as the bedrock for comprehending the functionality and optimization intricacies of deep learning models. Subsequently, I embarked on the development and meticulous evaluation of a sophisticated deep learning model, systematically assessing its performance across a spectrum of performance metrics, including accuracy, precision, recall, and the F1 score. These metrics collectively furnished a comprehensive appraisal of the model's predictive prowess and its adeptness in accurately discerning between genuine and fake Instagram accounts. Ultimately, our findings underscore the efficacy of the developed model in effectively identifying fake Instagram profiles, as evidenced by its remarkable accuracy rate of 95% on the test dataset. This achievement underscores the potential of our model in mitigating the pernicious impact of spam and fraudulent activities on social media platforms, thereby fostering a more secure and trustworthy online ecosystem for users worldwide.

6. Conclusion

The phenomenon of fake followers on social media platforms, particularly Twitter, represents a significant challenge as these accounts are created solely to artificially inflate the follower count of target accounts. Beyond being a concern for the platform itself, fake followers have far-reaching implications, influencing various aspects of economy, politics, and society by distorting popularity and influence metrics. As such, the primary objective of this bachelor's thesis was to develop and assess effective techniques for detecting fake Twitter followers, with the ultimate goal of mitigating their impact. The proliferation of fake followers not only undermines the credibility of social platforms but also poses risks to the integrity of online interactions and the reliability of information disseminated through these channels. Therefore, the need for robust and efficient detection methods is paramount in combating this deceptive practice.

Throughout this thesis, a comprehensive examination of various detection techniques and methodologies has been undertaken. By leveraging data analysis, machine learning algorithms, and social network analysis, I have explored innovative approaches to identify patterns and characteristics indicative of fake follower accounts. Through meticulous experimentation and evaluation, the performance of these techniques has been rigorously assessed, with a focus on accuracy, precision, recall, and computational efficiency. The findings of this research offer valuable insights into the effectiveness of different detection strategies and highlight the importance of feature engineering, model selection, and algorithm optimization in enhancing detection performance. By leveraging advanced techniques such as anomaly detection, clustering, and classification, I have demonstrated the feasibility of accurately identifying fake followers based on their behavioral attributes, engagement patterns, and network characteristics.

This thesis underscores the significance of collaboration between researchers, social media platforms, and regulatory bodies in addressing the challenge of fake followers. By sharing knowledge, data, and best practices, I can develop more robust detection algorithms and strategies that adapt to evolving tactics employed by malicious actors. From a practical standpoint, the outcomes of this research have implications for social media platforms, policymakers, and users alike. Social platforms must prioritize the development and implementation of automated detection mechanisms to mitigate the proliferation of fake

followers and safeguard user trust. Policymakers, on the other hand, can play a crucial role in establishing regulations and guidelines to combat deceptive practices and protect the integrity of online ecosystems.

While the study on fake profile detection on social networks presents valuable insights and contributes to the ongoing efforts in combating online deception, it is important to acknowledge certain shortcomings that warrant consideration. The study may be limited by the availability and quality of data used for training and evaluation purposes. Access to comprehensive and representative datasets of fake profiles can be challenging, leading to potential biases and inaccuracies in the performance assessment of detection techniques. Additionally, the reliance on publicly available data may not capture the full spectrum of sophisticated fake profiles that evade detection, thus limiting the generalizability of the findings. The evaluation metrics used to assess the performance of detection techniques may not fully capture the effectiveness of the methods in real-world scenarios. While metrics such as accuracy, precision, recall, and ROC AUC provide quantitative measures of performance, they may not fully reflect the practical utility and robustness of the algorithms in detecting evolving tactics employed by malicious actors.

The study may overlook the ethical implications associated with fake profile detection, including privacy concerns and unintended consequences. The deployment of detection techniques may inadvertently lead to the misclassification of genuine accounts as fake, resulting in user frustration and distrust. Moreover, the use of automated detection methods raises questions about algorithmic bias and fairness, particularly in cases where certain demographics or communities are disproportionately affected. It may lack comprehensive validation and reproducibility of results, which are essential for ensuring the reliability and validity of findings. Transparent reporting of methodologies, code, and data sources is crucial for enabling peer review, replication, and validation by other researchers, thereby strengthening the credibility and robustness of the study outcomes. While the study on fake profile detection on social networks offers valuable insights and advancements in the field, it is imperative to acknowledge and address the aforementioned shortcomings to ensure the effectiveness, fairness, and ethical integrity of detection techniques in practice. By addressing these limitations, future research endeavors can contribute to the development of more reliable, accurate, and ethical approaches for combating fake profiles and enhancing trust and security in online environments.

7. References

1. Adebowale, M.A., Lwin, K.T. and Hossain, M.A., 2023. Intelligent phishing detection scheme using deep learning algorithms. *Journal of Enterprise Information Management*, *36*(3), pp.747-766.
2. Hao, P. and Wang, X., 2019. Integrating PHY security into NDN-IoT networks by exploiting MEC: Authentication efficiency, robustness, and accuracy enhancement. *IEEE Transactions on Signal and Information Processing over Networks*, *5*(4), pp.792-806.
3. Fire, M., Goldschmidt, R. and Elovici, Y., 2014. Online social networks: threats and solutions. *IEEE Communications Surveys & Tutorials*, *16*(4), pp.2019-2036.
4. Al-Qurishi, M., Rahman, S.M.M., Hossain, M.S., Almogren, A., Alrubaian, M., Alamri, A., Al-Rakhami, M. and Gupta, B.B., 2018. An efficient key agreement protocol for Sybil-precaution in online social networks. *Future Generation Computer Systems*, *84*, pp.139-148.
5. Foody, M., Samara, M. and Carlbring, P., 2015. A review of cyberbullying and suggestions for online psychological therapy. *Internet Interventions*, *2*(3), pp.235-242.
6. statista, "Number of monthly active instagram users,"2020, last accessed on 2020-11-14. [Online]. Available: https://www.statista.com/statistics/253577/number-of-monthlyactive-instagram-users/
7. Sapountzi, A. and Psannis, K.E., 2018. Social networking data analysis tools & challenges. *Future Generation Computer Systems*, *86*, pp.893-913.
8. Zhang, Z. and Gupta, B.B., 2018. Social media security and trustworthiness: overview and new direction. *Future Generation Computer Systems*, *86*, pp.914-925.
9. Adewole, K.S., Han, T., Wu, W., Song, H. and Sangaiah, A.K., 2020. Twitter spam account detection based on clustering and classification methods. *The Journal of Supercomputing*, *76*, pp.4802-4837.
10. Arshad, H., Omlara, E., Abiodun, I.O. and Aminu, A., 2020. A semi-automated forensic investigation model for online social networks. *Computers & Security*, *97*, p.101946.

11. Rout, R.R., Lingam, G. and Somayajulu, D.V., 2020. Detection of malicious social bots using learning automata with url features in twitter network. *IEEE Transactions on Computational Social Systems*, *7*(4), pp.1004-1018.
12. Rodríguez-Ruiz, J., Mata-Sánchez, J.I., Monroy, R., Loyola-Gonzalez, O. and López-Cuevas, A., 2020. A one-class classification approach for bot detection on Twitter. *Computers & Security*, *91*, p.101715.
13. Latah, M., 2020. Detection of malicious social bots: A survey and a refined taxonomy. *Expert Systems with Applications*, *151*, p.113383.
14. Wanda, P. and Jie, H.J., 2020. DeepProfile: Finding fake profile in online social network using dynamic CNN. *Journal of Information Security and Applications*, *52*, p.102465.
15. Boshmaf, Y., Logothetis, D., Siganos, G., Lería, J., Lorenzo, J., Ripeanu, M., Beznosov, K. and Halawa, H., 2016. Íntegro: Leveraging victim prediction for robust fake account detection in large scale OSNs. *Computers & Security*, *61*, pp.142-168.
16. Muñoz, S.D. and Pinto, E.P.G., 2020, December. A dataset for the detection of fake profiles on social networking services. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 230-237). IEEE.
17. Quinlan, J.R., 1987. Simplifying decision trees. *International journal of man-machine studies*, *27*(3), pp.221-234.
18. Westreich, D., Lessler, J. and Funk, M.J., 2010. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of clinical epidemiology*, *63*(8), pp.826-833.
19. Breiman, L., 2001. Random forests. *Machine learning*, *45*, pp.5-32.
20. Freund, Y. and Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, *55*(1), pp.119-139.
21. Williams, C.K. and Barber, D., 1998. Bayesian classification with Gaussian processes. *IEEE Transactions on pattern analysis and machine intelligence*, *20*(12), pp.1342-1351.
22. Smola, A.J. and Schölkopf, B., 2004. A tutorial on support vector regression. *Statistics and computing*, *14*, pp.199-222.

23. Jeatrakul, P. and Wong, K.W., 2009, October. Comparing the performance of different neural networks for binary classification problems. In *2009 Eighth International Symposium on Natural Language Processing* (pp. 111-115). IEEE.
24. Bhumiratana, B., 2011, November. A model for automating persistent identity clone in online social network. In 2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (pp. 681-686). IEEE.
25. Toloşi, L. and Lengauer, T., 2011. Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics*, *27*(14), pp.1986-1994.
26. Wani, M.A., Jabin, S., Yazdani, G. and Ahmadd, N., 2018. Sneak into devil's colony-A study of fake profiles in online social networks and the cyber law. *arXiv preprint arXiv:1803.08810*.
27. Sharma, S.K. and Wang, X., 2019. Toward massive machine type communications in ultra-dense cellular IoT networks: Current issues and machine learning-assisted solutions. *IEEE Communications Surveys & Tutorials*, *22*(1), pp.426-471.
28. Sowmya, P. and Chatterjee, M., 2020, July. Detection of fake and clone accounts in twitter using classification and distance measure algorithms. In 2020 International Conference on Communication and Signal Processing (ICCSP) (pp. 0067-0070). IEEE.