e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 12, December 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

© 2024 IJMRSET | Volume 7, Issue 12, December 2024| DOI: 10.15680/IJMRSET.2024.0712007

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Detection of Fake and Clone Profiles in Online Social Networks Using Rule-Based and Machine Learning Techniques

Palagati Anusha[1], Adki Sai Vinay[2], Bhupalam Lakshmi Rohit[3], Budde Rachana[4]

Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Ibrahimpatnam, Telangana, India[1]

Student, Department of CSE, Guru Nanak Institute of Technology, Ibrahimpatnam, Telangana, India[2,3,4]

ABSTRACT: Online Social Networks (OSNs) have become platforms where people with shared interests or real-world connections interact. However, as OSNs grow in popularity, so do the security and privacy concerns. One significant threat is the creation of fake and cloned profiles. Cloning involves stealing a user's details to create a duplicate profile, which can be misused to damage the original profile owner's identity and carry out malicious activities such as phishing, stalking, or spamming. A fake profile, on the other hand, is one that impersonates a person or organization that does not exist, typically for fraudulent purposes.

This paper proposes a method to detect fake and cloned profiles specifically on Twitter. The detection of fake profiles is carried out using a set of rules that effectively distinguish between genuine and fake profiles. For detecting cloned profiles, two methods are employed: one based on similarity measures and the other utilizing the C4.5 decision tree algorithm. In the similarity-based approach, two types of similarities are considered—similarity of attributes and similarity of network relationships. The C4.5 algorithm, on the other hand, detects clones by constructing a decision tree based on information gain. The paper also compares the effectiveness of these two methods in identifying cloned profiles.

KEYWORDS: Online Social Networks (OSN), Fake Profiles, Profile Cloning, Twitter Security, Similarity Measures, C4.5 Decision Tree Algorithm, Identity Theft Detection, Social Network Privacy.

I. INTRODUCTION

NLINE Social Networks (OSN) like Facebook, Twitter, LinkedIn, Instagram etc are used by billions of users all around the world to build network connections. The ease and accessibility of social networks have created a new era of networking. OSN users share a lot of information in the network like photos, videos, school name, college name, phone numbers, email address, home address, family relations, bank details, career details etc. This information if put into hands of attackers, the after effects are very severe. Most of the OSN users are unaware of the security threats that exist in Sowmya P is with the Department of Computer Engineering, Pillai College of Engineering, University of Mumbai, Maharashtra, India

Madhumita Chatterjee is with the Department of Computer Engineering, Pillai HOC College of Engineering and Technology, University of Mumbai, Maharashtra, India . the social networks and easily fall prey to these attacks. The risks are more dangerous if the victims are children. In Profile Cloning attack, the profile information of existing users are stolen to create duplicate profiles and these profiles are misused for spoiling the identity of original profile owners . There are two types of Profile Cloning namely - Same Site and Cross Site Profile Cloning.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

If user credentials are taken from one Network to create a clone profile in same Network then it is called Same Site profile cloning. In Cross Site profile cloning, attacker takes the user information from one Network to create a duplicate profile in other Network in which the user is not having any account.

As the registration process in social networks have become very simple in order to attract more and more users, the creation of fake profiles are also increasing in an alarming rate. An attacker creates a fake profile in order to connect to a victim to cause malicious activities. And also to spread fake news and spam messages.

What Is A Social Network?

Wikipedia defines a social network service as a service which "focuses on the building and verifying of online social networks for communities of people who share interests and activities, or who are interested in exploring the interests and activities of others, and which necessitates the use of software."

A report published by OCLC provides the following definition of social networking sites: "Web sites primarily designed to facilitate interaction between users who share interests, attitudes and activities.

Opportunities and Challenges

The popularity and ease of use of social networking services have excited institutions with their potential in a variety of areas. However effective use of social networking services poses a number of challenges for institutions including long-term sustainability of the services; user concerns over use of social tools in a work or study context; a variety of technical issues and legal issues such as copyright, privacy, accessibility; etc. Institutions would be advised to consider carefully the implications before promoting significant use of such service

II. LITERATURE SURVEY

S. Wen et. Al (2018) In this Paper Social network worms, such as email worms and facebook worms, pose a critical security threat to the Internet. Modeling their propagation dynamics is essential to predict their potential damages and develop countermeasures. Although several analytical models have been proposed for modeling propagation dynamics of social network worms, there are two critical problems unsolved: temporal dynamics and spatial dependence. First, previous models have not taken into account the different time periods of Internet users checking emails or social messages, namely, temporal dynamics. Second, the problem of spatial dependence results from the improper assumption that the states of neighboring nodes are independent. These two problems seriously affect the accuracy of the previous analytical models. To address these two problems, we propose a novel analytical model. This model implements a spatial-temporal synchronization process, which is able to capture the temporal dynamics. Additionally, we find the essence of spatial dependence is the spreading cycles. By eliminating the effect of these cycles, our model overcomes the computational challenge of spatial dependence and provides a stronger approximation to the propagation dynamics. To evaluate our susceptible-infectious-immunized (SII) model, we conduct both theoretical analysis and extensive simulations. Compared with previous epidemic models and the spatial-temporal model, the experimental results show our SII model achieves a greater accuracy. We also compare our model with the susceptible-infectious-susceptible and susceptible-infectious-recovered models. The results show that our model is more suitable for modeling the propagation of social network worms.

Sentiment Analysis of Social Media Networking Sites: A Comparative Study on User Engagement and Public Opinion Authors

Mr. A Sandeep Reddy and Chokkamreddy Prakash (2024) Social media platforms such as Twitter, Instagram, and Facebook play a crucial role in shaping public opinion, consumer behavior, and user engagement. This study investigates the impact of sentiment on user engagement and loyalty across these platforms. Employing sentiment analysis techniques, the research examines the distribution of positive, neutral, and negative sentiments in user posts and their relationship with engagement metrics such as likes, retweets, and shares. The findings reveal that positive sentiment significantly enhances user engagement and loyalty, with Instagram showing the highest proportion of positive sentiment and lowest of negative sentiment. Comparative analysis across platforms indicates variations in sentiment distribution, with Instagram leading in positive interactions, while Twitter and Facebook exhibit more diverse sentiment profiles. The results highlight the importance of positive sentiment in driving user interactions and loyalty, offering valuable insights for social media marketers and content creators to optimize their strategies. The study

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

contributes to a deeper understanding of sentiment dynamics in social media and provides actionable recommendations for enhancing user engagement and brand perception.

E. Lebensztayn et. al(2011) In this study propose a realistic generalization of the Maki–Thompson rumour model by assuming that each spreader ceases to propagate the rumour right after being involved in a random number of stifling experiences. We consider the process with a general initial configuration and establish the asymptotic behaviour (and its fluctuation) of the ultimate proportion of ignorants as the population size grows to ∞. Our approach leads to explicit formulas so that the limiting proportion of ignorants and its variance can be computed.

L. Li et. al(2020) During the ongoing outbreak of coronavirus disease (COVID-19), people use social media to acquire and exchange various types of information at a historic and unprecedented scale. Only the situational information are valuable for the public and authorities to response to the epidemic. Therefore, it is important to identify such situational information and to understand how it is being propagated on social media, so that appropriate information publishing strategies can be informed for the COVID-19 epidemic. This article sought to fill this gap by harnessing Weibo data and natural language processing techniques to classify the COVID-19-related information into seven types of situational information. We found specific features in predicting the reposted amount of each type of information. The results provide data-driven insights into the information need and public attention.

S. Sommariva et. al(2018) Background: The importance of social networking sites (SNSs) as platforms to engage in the correction of "fake news" has been documented widely. More evidence is needed to understand the popularity of health-related rumors and how Health Educators can optimize their use of SNSs. Purpose: The purpose of this study was to explore the spread of health rumors and verified information on SNSs using the Zika virus as a case study. Methods: A content analysis of Zika-related news stories on SNSs between February 2016 and January 2017 was conducted to verify accuracy (phase 1). Phase 1 was followed by an analysis of volume of shares (phase 2) and a thematic analysis of headlines (phase 3). Results: Rumors had three times more shares than verified stories. Popular rumors portray Zika as a conspiracy against the public and a low-risk issue and connect it to the use of pesticides. Discussion: This study identifies the value of integrating in-depth analysis of popular health-related rumors into the development of communication strategies. Translation to Health Education Practice: Misinformation on SNSs can hinder disease prevention efforts. This study shows how information circulating on SNSs can be analyzed from a quantitative and qualitative standpoint to help Health Educators maximize the use of online communication platforms.

Y. Xiao et. al(2019) In the online social network, the spreading process of rumor contains complex dynamics. The traditional research of the rumor propagation mainly studies the spreading process of rumor from the perspectives of rumor and participating user. The symbiosis and confrontation of rumor and anti-rumor information and the dynamic changes of the influence of anti-rumor information are not emphasized. At the same time, people's profitability and herd psychology are also ignored. In view of the above problems, we fully consider the anti-rumor information and user's psychological factors, construct a rumor propagation dynamics model based on evolutionary game and anti-rumor information, and provide a theoretical basis for studying the inherent laws in the spreading process of rumor. First of all, we analyze the interaction pattern and characteristic of rumor in social network. In allusion to the symbiosis of rumor and anti-rumor information and the dynamic changes of the influence of anti-rumor information, we constructed the SKIR rumor propagation model based on the SIR model. Secondly, due to rivalry between rumor and anti-rumor information, as well as the user's profitability and herd psychology, we use evolutionary game theory to construct the driving force mechanism of information and explore the causes of user behavior in the spreading process of rumor. At the same time, we combine the behavior factors and external factors of the user to build the influence of information by multivariate linear regression method, which provides the theoretical basis for the driving force of information. Finally, combining the SKIR model proposed in this paper, we get a rumor propagation dynamics model based on evolutionary game and anti-rumor information. We have proved by experiments that the model can effectively describe the propagating situation of rumor and the dynamic change rule of the influence of anti-rumor information. On the other hand, it can also reflect the influence of people's psychology on rumor propagation.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Existing System

- Brodka, Mateusz Sobas and Henric Johnson in their paper have proposed two novel methods for detecting cloned profiles. The first method is based on the similarity of attribute values from original and cloned profiles and the second method is based on the network relationships. A person who doubts that his profile has been cloned will be chosen as a victim. Then treating name as primary key, a search is made for profiles with the same name as that of victim, using query search. Potential clone (Pc) and the Victim profile (Pv) are compared and similarity S is calculated. If S(Pc, Pv) > Threshold, then profile is suspected to be a clone. In the verification step, the user does it manually as he knows which is his original profile and which one is a duplicate.
- Cresci S, Di Pietro R, Petrocchi M, Spognardi A, Tesconi M, in their paper have reviewed some of the most relevant existing features and rules (proposed by Academia and Media) for fake Twitter accounts detection. They have used these rules and features to train a set of machine learning classifiers. Then they have come up with Class A classifier which can effectively classify original and fake accounts.

Existing System Disadvantages

- Fake and clone profiles have become a very serious social threat. As information like phone number, email id, school or college name, company name, location etc are readily exposed in social networks, hackers can easily hack this information to create fake or clone profiles.
- They then try to cause various attacks like phishing, spamming, cyberbullying etc. They even try to defame the legitimate owner or the organisation.

Proposed System

- In the proposed system we used to detect fake Twitter profiles. Here fake profiles are detected based on rules that effectively distinguish fake profiles from genuine ones. Some of the rules that are used to detect fake profiles are - usually fake profiles do not have profile name or image. They do not include any description about the account. The geo-enabled field will be false as they do not want to expose their location in tweets. They usually make large number of tweets or sometimes the profiles would not have made any tweets etc. The rules are applied on the profile, for each matching rule, a counter is incremented, if the counter value is greater than pre-defined threshold, then the profile is termed as fake.

Proposed System Advantages

- The modules worked fine and was able to detect clones with good accuracy.
- Good Results

System Architecture

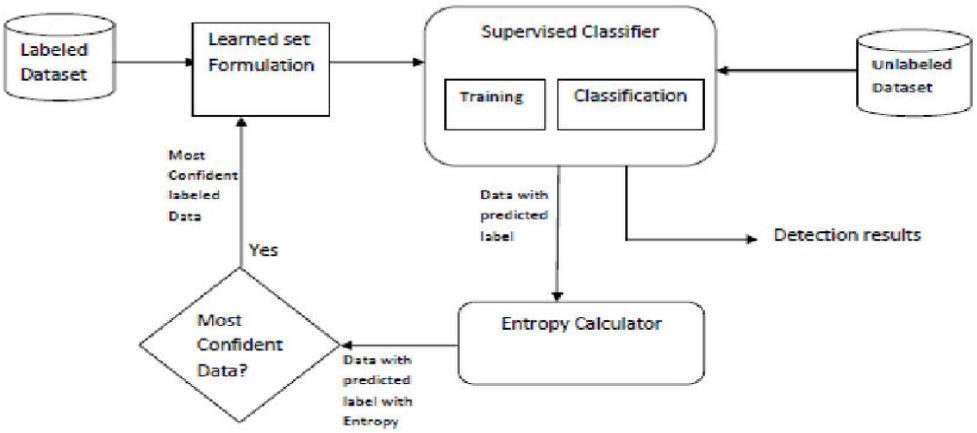


Fig no:-1 System Architecture

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Explanation

This system architecture diagram presents a methodology for detecting fake profiles in social media using a supervised classification approach. Here's a breakdown of each component and its role in the detection process:

- **Labeled Dataset:** This is a collection of data samples where each sample is labeled (marked as either fake or genuine). This labeled dataset provides a base for training the classifier, allowing it to learn how to distinguish between fake and real profiles based on various features.
- **Learned Set Formulation:** This module takes the labeled data and processes it to form a "learned set." This set serves as the foundation for training a supervised classifier, which will use this knowledge to classify new, unseen data samples.
- **Supervised Classifier:** The supervised classifier is the core of the system. It has two main stages
- **Training:** The classifier is trained on the labeled dataset, where it learns patterns and distinctions between fake and genuine profiles.
- **Classification:** Once trained, the classifier can then analyze new data (unlabeled dataset) and assign labels, predicting whether a profile is fake or genuine.
- **Unlabeled Dataset:** This contains data samples that haven't been labeled yet. The classifier processes this data to make predictions on whether each sample represents a fake or genuine profile.
- **Entropy Calculator:** After classification, the system sends the labeled predictions to the entropy calculator. This module assesses the "confidence" or "uncertainty" of the predictions made by the classifier. Entropy is a measure of uncertainty, with higher entropy indicating less confidence in the prediction. This calculation helps identify which predictions are the most confident and which require further review.
- **Decision Check (Most Confident Data?):** This decision block examines whether the labeled data predictions are "confident" enough. If the classifier is confident in its predictions (low entropy), those predictions are added to the "learned set" for continuous learning. If not, they may require more scrutiny or additional data before a decision can be made.
- **Detection Results**: The final output is a set of labeled predictions for the unlabeled dataset, representing whether each profile is detected as fake or genuine. These detection results can then be used for further action, like flagging fake profiles for review.

III. METHODOLOGIES

Modules

Data Collection:

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform.

Dataset:

The dataset consists of 1338 individual data. There are 9 columns in the dataset, which are described below.
ID: Id number
UserID: twitter id
No of Abuse Report: The number of Abuse Report
No of Rejected Friend Requests: The number of Rejected Friend Requests Followers in the twitter amount
No of Friends: The number of people friends in the twitter amount
No of Followers: The number of people Followers in the twitter amount
No of Likes To Unknown Account: The number of Likes To Unknown Account
No of Comments Per Day: The number of Comments Per Day
Fake or Not Category: 1 OR 0

Data Preparation:

we will transform the data. By getting rid of missing data and removing some columns. First we will create a list of column names that we want to keep or retain.Next we drop or remove all columns except for the columns that we want to retain.Finally we drop or remove the rows that have missing values from the data set.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Model Selection:
While creating a machine learning model, we need two dataset, one for training and other for testing. But now we have only one. So lets split this in two with a ratio of 80:20. We will also divide the dataframe into feature column and label column. Here we imported train_test_split function of sklearn. Then use it to split the dataset. Also, *test_size = 0.2*, it makes the split with 80% as train dataset and 20% as test dataset. The *random_state* parameter seeds random number generator that helps to split the dataset. The function returns four datasets. Labelled them as *train_x, train_y, test_x, test_y.* If we see shape of this datasets we can see the split of dataset. We will use Random Forest Classifier, which fits multiple decision tree to the data. Finally I train the model by passing *train_x, train_y* to the *fit* method. Once the model is trained, we need to Test the model. For that we will pass *test_x* to the predict method.

Random Forest is one of the most powerful methods that is used in machine learning for classification problems. The random forest comes in the category of the supervised regressor algorithm. This algorithm is carried out in two different stages the first one deals with the creation of the forest of the given dataset, and the other one deals with the prediction from the regressor.

Analyze and Prediction:
In the actual dataset, we chose only 7 features :
UserID: twitter id
No of Abuse Report: The number of Abuse Report
No of Rejected Friend Requests: The number of Rejected Friend Requests Followers in the twitter amount
No of Friends: The number of people friends in the twitter amount
No of Followers: The number of people Followers in the twitter amount
No of Likes To Unknown Account: The number of Likes To Unknown Account
No of Comments Per Day: The number of Comments Per Day
Accuracy on test set:
We got a accuracy of 95.1% on test set.

Saving the Trained Model:
Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle .Make sure you have pickle installed in your environment.Next, let's import the module and dump the model into .pkl file.

Implementation
Existing algorithms, such as those relying on similarity measures, have been widely used to detect cloned profiles. These methods often compare attributes like names, follower counts, or network relationships to determine potential clones. While effective to an extent, these algorithms face limitations, particularly when the data lacks distinct patterns or includes noise. Another common issue is the manual verification step often required in similarity-based methods, which makes the process time-consuming and less scalable for platforms with millions of users.

To address these challenges, we adopted the C4.5 decision tree algorithm, a machine learning approach known for its ability to handle complex and noisy datasets effectively. The primary reason for choosing C4.5 lies in its ability to make decisions based on a clear, hierarchical structure derived from the data itself. Unlike similarity measures, which may fail when patterns are subtle or attributes overlap, C4.5 leverages information gain to determine the most impactful attributes for classification. This ensures that the algorithm focuses on the features that best distinguish fake profiles from genuine ones, even in complex datasets.

C4.5 builds a decision tree by splitting the dataset iteratively, using the attribute that provides the highest information gain at each step. For Twitter, the dataset includes attributes such as the number of abuse reports, rejected friend requests, followers, friends, likes to unknown accounts, and comments per day. The algorithm identifies patterns among these attributes, such as unusually high abuse reports or low engagement rates, which are common indicators of fake profiles. By applying these patterns, the decision tree classifies profiles as either fake or genuine in a systematic and automated manner. What makes C4.5 particularly effective is its handling of real-world complexities. It can process both continuous and categorical data, enabling the algorithm to deal with diverse types of user behavior on social

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

media. For example, continuous attributes like "comments per day" are split using thresholds determined during training, while categorical attributes such as "geo-enabled status" are used directly. Additionally, the algorithm is resilient to missing data, estimating missing values using probabilities to maintain accuracy. This robustness ensures consistent performance even when the data is incomplete or noisy, a common scenario in social media datasets.The decision tree generated by C4.5 is interpretable and intuitive, making it easy to understand the classification process. For instance, a profile might be classified as fake if it has more than a certain number of abuse reports and lacks a profile picture, while another might be identified as genuine if it has a verified badge and a high friend-to-follower ratio. These clear decision paths help ensure the reliability of the classification and provide actionable insights for platform administrators.By adopting the C4.5 decision tree algorithm, we achieved significant improvements in detecting fake profiles. The structured decision-making process not only addresses the shortcomings of previous approaches but also provides scalability and adaptability to evolving patterns of fake accounts. The algorithm's ability to learn from labeled data and generalize to new, unseen data makes it a robust solution for enhancing security on platforms like Twitter.

Experimental Results

Fig no:-2 Home Page

This is the home page for fake user identification on social network. It consists of navigation to abstract for the project, upload data set, prediction history, analysis tab and also includes future enhancements for this project. Through this page we can able to navigate to login page.

Fig no:-3 Login Page

© 2024 IJMRSET | Volume 7, Issue 12, December 2024| DOI: 10.15680/IJMRSET.2024.0712007

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A login page for users who are trying to predicts fake users, to access the resources or system usually require username and password.

Fig no:-4 Login Entry

Here we enter the login details to access the system resources

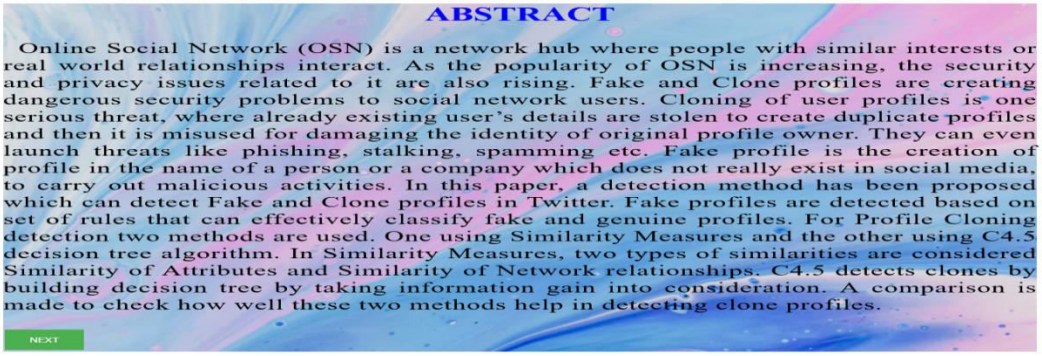


Fig no:-5 Abstract Page

After entering the login page details correctly then we see an abstract page pop up where it describes in brief about what is the project and the purpose of and also what algorithm is being used.

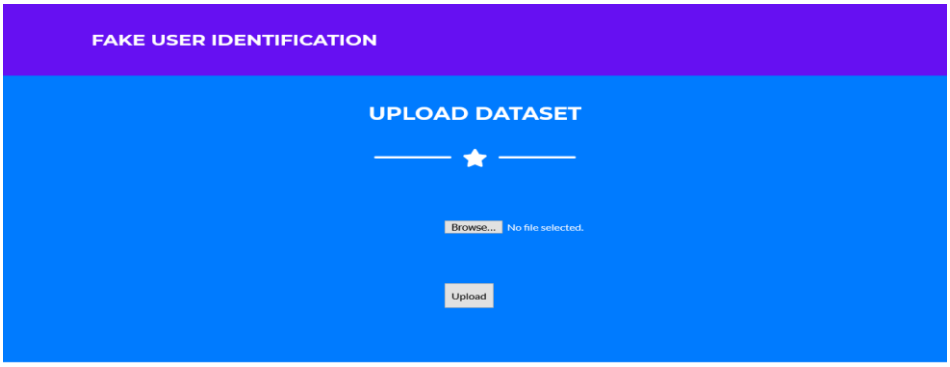


Fig no:-6 Upload Dataset Page

© 2024 IJMRSET | Volume 7, Issue 12, December 2024| DOI: 10.15680/IJMRSET.2024.0712007

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

After we click next in the abstract page then we see page for uploading data set where we need to provide the dataset to train the model.

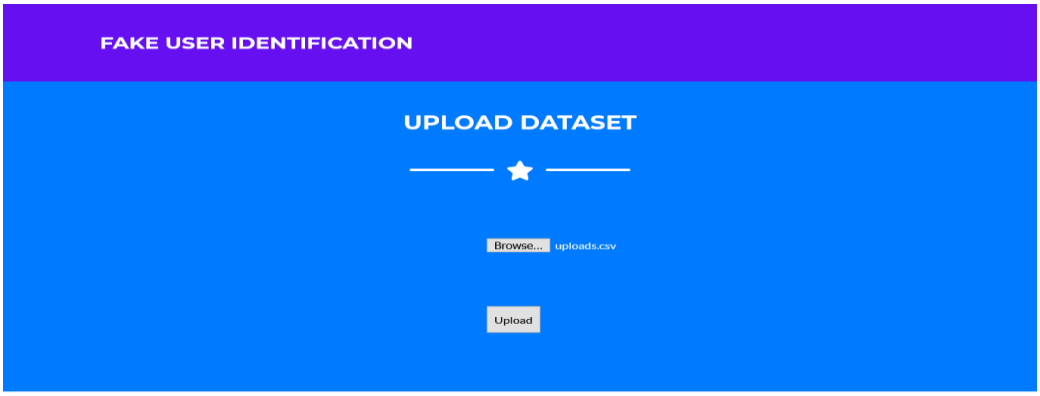


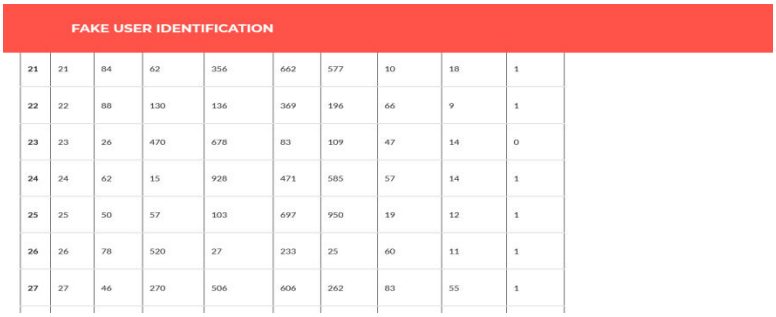
Fig no:-7 Data set uploading page

Here we uploaded the dataset and clicked on upload button

	UserID	No Of Abuse Report	No Of Rejected Friend Requests	No Of Freind Requests Thar Are Not Accepted	No Of Friends	No Of Followers	No Of Likes To Unknown Account	No Of Comments Per Day	Fake Or Not Category
Id									
1	1	37	415	204	290	838	26	53	1

Fig no:-8 Data set preview page

We upload a CSV file and this preview page displays our uploaded data in the tabular form. It consists of UserId, no of abuse report, no of rejected friend request, no of friend request that are not accepted, no of friends, no of followers , no of likes to unknown account, no of comments per day and the classification of data.



21	21	84	62	356	662	577	10	18	1
22	22	88	130	136	369	196	66	9	1
23	23	26	470	678	83	109	47	14	0
24	24	62	15	928	471	585	57	14	1
25	25	50	57	103	697	950	19	12	1
26	26	78	520	27	233	25	60	11	1
27	27	46	270	506	606	262	83	55	1

Fig no:-9 Data set preview page

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

We upload a CSV file and this preview page displays our uploaded data in the tabular form. It consists of UserId, no. of abuse report, no. of rejected friend request, no. of friend request that are not accepted, no of friends, no. of followers , no. of likes to unknown account, no. of comments per day and the classification of data.

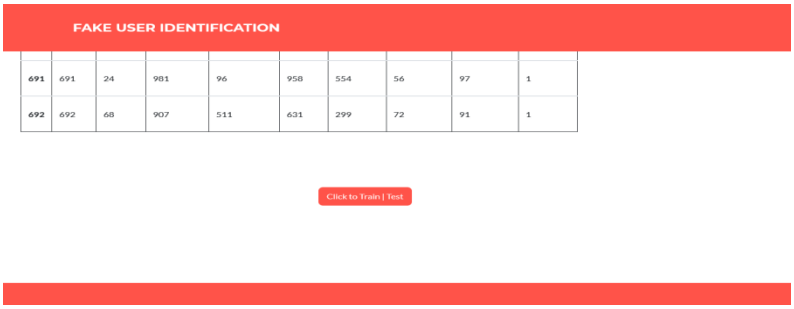


Fig no:-10 Dataset preview

After previewing the page and checking whether the data set has been uploaded properly then click train test button to train the model.

Fig no:-11 Prediction page

After the training of the model is finished, then it displays the prediction page where we can enter user id and other details to check whether the profile with the information that we have entered is a genuine profile or a fake profile.

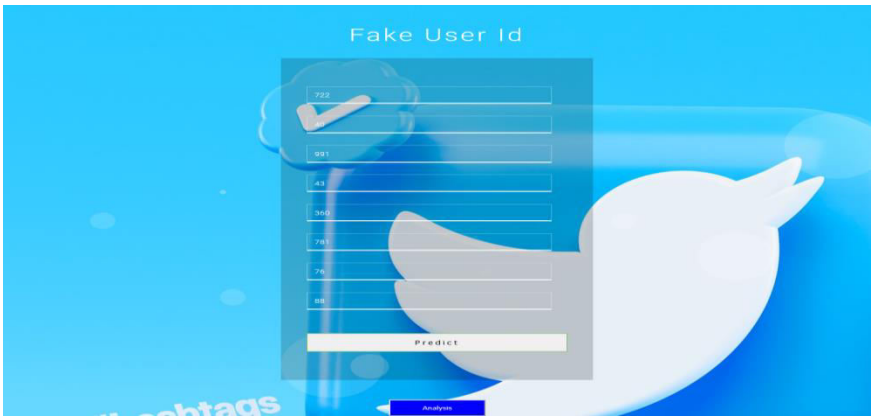


Fig no:-12 User information entry page

© 2024 IJMRSET | Volume 7, Issue 12, December 2024| DOI: 10.15680/IJMRSET.2024.0712007

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

In this we entered the information that we want to check whether it is genuine or not.

Fig no:-13 Result display

After we click on predict button it analysis the data based on our trained model and displays the result whether the profile that we wanted to check is genuine profile or fake profile.

Fig no:-14 User information entry page

In this we entered the information that we want to check whether it is genuine or not.

Fig no:-15 Result display

© 2024 IJMRSET | Volume 7, Issue 12, December 2024| DOI: 10.15680/IJMRSET.2024.0712007

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

After we click on predict button it analysis the data based on our trained model and displays the result whether the profile that we wanted to check is genuine profile or fake profile.

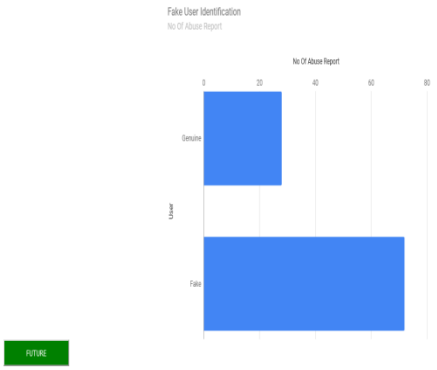


Fig no:-16 Analysis Page

After we are displayed with the result we can click on the analysis button where we are displayed with analysis page representing the information about how many users reported this profile as a genuine profile or a fake profile in the form of a bar graph where it shows the count of people reported it as fake or genuine.

IV. CONCLUSION

Fake and clone profiles have become a very serious problem in online social networks. We hear some or the other threats caused by these profiles in everyday life. So a detection method has been proposed which can find both fake and clone Twitter profiles. For fake detection, a set of rules were used which when applied can classify fake and genuine profiles. Clone detection was carried out using Similarity Measures and C4.5 algorithm and a comparison was made to check the performance. Clone detection using Similarity Measures worked better than C4.5 and was able to detect most of the clones which were fed into the system.

V. FUTURE ENHANCEMENT

In this work, we have focused on identifying fake users in social networks using the C4.5 decision tree algorithm by relying primarily on profile attributes for classification. However, the detection process can be significantly enhanced in the future by incorporating additional data sources and advanced methodologies. For instance, analyzing user-generated content such as posts, tweets, and comments using Natural Language Processing (NLP) can provide deeper insights into user authenticity. Behavioral patterns such as posting frequency, interaction types, and network structure could be integrated to identify anomalies indicative of fake accounts.

REFERENCES

1. Sowmya P and Madhumita Chatterjee ,” Detection of Fake and Cloned Profiles in Online Social Networks”, Proceedings 2019: Conference on Technologies for Future Cities (CTFC)
2. Georgios Kontaxis, Iasonas Polakis, Sotiris Ioannidis and Evangelos P. Markatos, “Detecting Social Network Profile Cloning”, 2013
3. Piotr Bródka, Mateusz Sobas and Henric Johnson, “Profile Cloning Detection in Social Networks”, 2014 European Network Intelligence Conference
4. Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angello Spognardi, Maurizio Tesconi, “Fame for sale: Efficient detection of fake Twitter followers”, 2015 Elsevier’s journal Decision Support Systems, Volume 80

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

5. Ahmed El Azab, Amira M Idrees, Mahmoud A Mahmoud, Hesham Hefny, "Fake Account Detection in Twitter Based on Minimum Weighted Feature set", World Academy of Science, Engineering and Technology, International Journal of Computer and Information Engineering Vol:10, 2016
6. M.A.Devmane and N.K.Rana, "Detection and Prevention of Profile Cloning in Online Social Networks", 2014 IEEE International Conference on Recent Advances and Innovations in Engineering
7. Kiruthiga. S, Kola Sujatha. P and Kannan. A, "Detecting Cloning Attack in Social Networks Using Classification and Clustering Techniques" 2014 International Conference on Recent Trends in Information Technology
8. Buket Erşahin, Ozlem Aktaş, Deniz Kilinç, Ceyhun Akyol, "Twitter fake account detection", 2017 International Conference on Computer Science and Engineering (UBMK)
9. Arpitha D, Shrilakshmi Prasad, Prakruthi S, Raghuram A.S, "Python based Machine Learning for Profile Matching", International Research Journal of Engineering and Technology (IRJET), 2018
10. Olga Peled, Michael Fire, Lior Rokach, Yuval Elovici, "Entity Matching in Online Social Networks", 2013 International Conference on Social Computing
11. Aditi Gupta and Rishabh Kaushal, "Towards Detecting Fake User Accounts in Facebook", 2017 ISEA Asia Security and Privacy (ISEASP)
12. Michael Fire, Roy Goldschmidt, Yuval Elovici, "Online Social Networks: Threats and Solutions", JOURNAL OF LATEX CLASS FILES, VOL. 11, NO. 4, DECEMBER 2012, IEEE Communications Surveys & Tutorials
13. Ashraf Khalil, Hassan Hajjdiab and Nabeel Al-Qirim, "Detecting Fake Followers in Twitter: A Machine Learning Approach" 2017 International Journal of Machine Learning and Computing
14. Mohammad Reza Khayyambashi and Fatemeh Salehi Rizi, "An approach for detecting profile cloning in online social networks" 2013 International Conference on e-Commerce in Developing Countries: with focus on e-Security
15. Mauro Conti, Radha Poovendran and Marco Secchiero, "FakeBook: Detecting Fake Profiles in On-line Social Networks", 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining
16. Reddy, M. A. S., & Prakash, C. (2024). Sentiment Analysis of Social Media Networking Sites: A Comparative Study on User Engagement and Public Opinion.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com