International Journal of Recent Advances in Multidisciplinary Topics
Volume 3, Issue 3, March 2022
https://www.ijramt.com | ISSN (Online): 2582-7839

Social Networks Fake Detection

Atul Prakash Patil[1*], Vitthal Remulkar[2], Hardik[3], Ulka Shirole[4]
[1,2,3]*Student, Department of Information Technology, A.C Patil College of Engineering, Navi Mumbai, India*
[4]*Guide and Professor, Department of Information Technology, A.C Patil College of Engineering, Navi Mumbai, India*

***Abstract*: Nowadays, social platforms have become part of our day by days life, we are more associated with these platforms, we use OSNs to stay connected to friends, share news, and we also use social media platforms to show our talent, run our e-business. This all-social media platform generates a large amount of personal data this all-personal data attracts attackers and imposters to steal personal data, share fake news, and perform malware attacks. By creating Human-generated or bot-generated fake profiles on social media, Therefore, in this project, we create a detection model, which detects fake profiles and genuine profiles on Twitter, based on visible features like followers count, friends count, status counts, and more by using various machine learning methods.**

***Keywords*: Fake profile, Bot, Twitter data, social media**

1. Introduction

in this project we are focusing on detecting fake profiles and smart BOTS on a social media platform like Twitter. Nowadays more than fake profile bots are used because it is automated and can we have operated without a human. bots and fake profiles generated for stilling personal data of users on social media platforms like Twitter also for spreading fake news and rumors that can perform a big impact on society as we go forward in technology, A. I, is now used in every field of work and taking place of humans, and now to detect bots is more serious than human-made fake profile. So, we create a model that detects the smart bots and human-generated fake profiles based on the Twitter data parameters like followers, tweets, following, etc. We are using the Twitter dataset for our model as we can fetch the real-time Twitter data of a user via Twitter API.

2. Literature Survey

A.[1] Vijay Tiwari, Ministry of Defense proposed Analysis and Detection of Fake Profile Over Social Network: In this paper on bot detection is done using three methods bot detection based on scrutiny of content, detection based on network graph and combination or hybrid approach Machine learning methods used in this paper was post method, accuracy detection models and supervised learning.

B.[2] in this paper Detecting Fake Account on social media: In this paper neural networks and support vector machine are used and within its algorithm data pre-processing is done after that in feature reduction principal component analysis, Spearman's Rank-Order Correlation, Relevance and Redundancy Analysis Technique, Markov Blanket Technique, Wrapper Feature Selection using SVM are used and SVM-NN classification are done.

C.[3] in this statement Facebook said that there almost 4.3 percent of its active user accounts are duplicate, and almost 83 million user accounts are fake which is increasing extremely fast.

D.[4] In this paper discus that social media growing extremely fast in field of entertainment, and business we use social media in every field but this all-social media platform having some issue like trolling, bulling, fraud mostly this done by using fake accounts

3. Existing System

In the current existing system, the machine learning algorithm used supervised learning machine learning algorithms is Random Forest, Decision Tree, and Naïve Bayes. These algorithms used has good accuracy but the model is incompatible to detect the fake profile in real time. It only allows the user to test on only selected dataset. The dataset used is of Twitter. Also, it only detected the fake profile but there are even non-human accounts or BOT on different social media that can also use to havoc and spam like many social evil/negative BOTS try to spread misinformation in public like in U.S election 2015. There is even good BOT that are made for good purpose like posting health tips, news alert, match score, etc. So, it fails to detect the positive BOTS and negative BOTS.

4. Proposed System

Before giving data to the model, we need to clean the data to increase accuracies i.e., remove unwanted columns, remove incorrect data values, remove absent values, etc. This process is known as the cleaning of data. This model classifies a profile as real or fake based on visible properties. The following characteristics are picked: friends, followers, status count, listed count, favorite count, geo-enabled, language Blank entries or NAN values are substituted by zeros.

5. System Architecture

1) Training of data

As shown in figure (1) there is a database that consists of all the Social Media Profiles which we have considered. This module is trained repetitively to obtain maximum accuracy using a classification algorithm. If a new profile is given to the

*Corresponding author: atulpatil7715@gmail.com

module, it will classify whether the given profile is fake or not. And then provides the appropriate result.

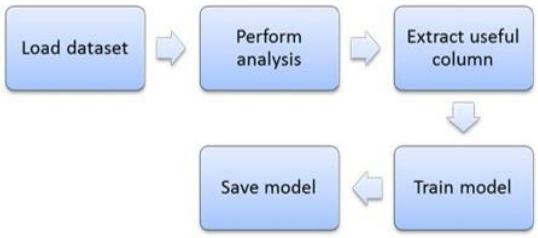

Fig. 1. Training of Dataset

2) Use case diagram

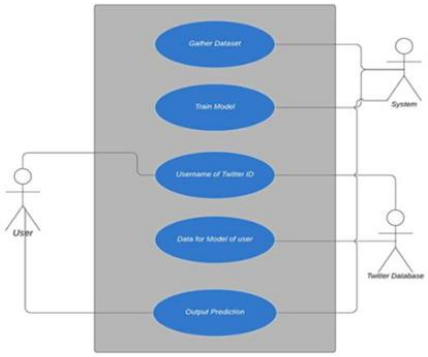

Fig. 2. Use Case Diagram

As shown in figure 2. When we give a new input profile first the module extracts the profile features. Then it goes through the Prediction model. then it will compare the features with the already trained dataset. Then the module will predict whether the profile is fake or not. It shows the output which is predetermined and trained. Then the output will be displayed as profile fake or not

3) Fake profile classification

Fake profile classification is the process of labeling an already identified profile as positive, negative, or neutral based on a variety of characteristics like followers, following, likes, tweets, and more. Here it is more crucial than we can distinguish negative from positive and neutral tweets.

4) Machine Learning

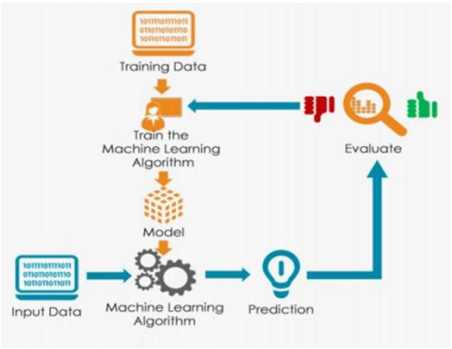

Fig. 3. Machine Learning

Machine learning is a method in which we train programming models based on ml algorithms with datasets we train a model using supervised learning, based on the Twitter dataset it will predict fake and genuine profiles. It is a specific subset of ai that trains machines how to learn, artificial intelligence based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention.

5) Supervised Learning

Supervised learning is the process of an algorithm learning from the training dataset to determine the mapping function from the input to the output. These problems can be further broken down into classification and regression.

6) Unsupervised Learning

Unsupervised learning is the process of an algorithm trying to model the underlying structure or distribution in the data to learn more about it. These problems are more complex than supervised learning and can be further grouped into clustering and association. Some popular examples are the aprioristic algorithm, k-means

6. Design Model

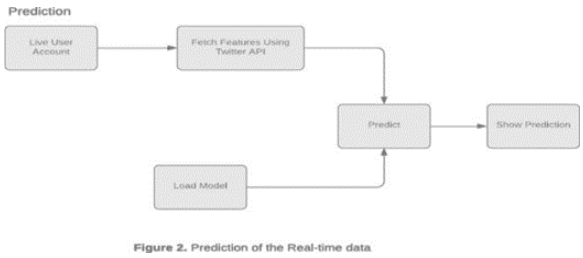

Fig. 4. Prediction Model

We are performing this with live user accounts. We will fetch features of a live user account using Twitter API. The model which we have built will be given to prediction and the fetched data will also be given to prediction. After this, it will give us a prediction of whether the user is fake or not.

7. Conclusion

In this project we have implemented a model which will detect fake profiles on social media platforms like Twitter, this model is user friendly, so users can easily interact, this model helps users to detect fake profiles or bots on social media that will help the user from malicious Attack on social media, for detecting fake profile we used classification algorithms in which random forest has more accuracy in large datasets

8. Future Scope

We have used a small twitter dataset to train the classifier, in future we can implement it on a large dataset and, we can implement it on mobile applications and desktops separately, so users can use this system on a mobile phone more efficiently for detecting fake profile.

References

[1] Tiwari, V. (2017). Analysis and detection of fake profile over social network. 2017 International Conference on Computing, Communication and Automation (ICCCA).

[2] Political advertising spending on Facebook between 2014 and 2018. Internet draft.
[3] CNBC. Facebook shares drop on news of fake accounts. Internet draft. 2012.
[4] Khaled, S., El-Tazi, N., & Mokhtar, H. M. O. (2018). Detecting Fake Accounts on social media. 2018 IEEE International Conference on Big Data (Big Data)
[5] R. Nithin Reddy &Nitesh Kumar, "Automatic Detection of Fake Profiles in Online Social Networks, "Computer Science and Engineering," National Institute of Technology, Rourkela.
[6] J R. Douceur, "The sybil attack," in international workshop on peer-to-peer systems. Springer, 2002, pp. 251–260.
[7] R. Kaur and S. Singh, "A survey of data mining and social network analysis-based anomaly detection techniques," Egyptian informatics journal, vol. 17, no. 2, pp. 199–216, 2016.
[8] L. M. Potgieter and R. Naidoo, "Factors explaining user loyalty in a social media-based brand community," Journal of Information Management, vol. 19, no. 1, pp. 1–9, 2017.
[9] Matt. Social Media Comparison Infographic.2014.
[10] Tumblr. URL: www.tumblr.com/.
[11] Foursquare. URL:https://foursquare.com/.