FEATURE BASED ANALYSIS ON THE DUPLICATE FAKE ACCOUNT ON TWITTER USING INTELLIGENT MECHANISM

Dr. Mohammed Ali Alhariri

College of Computing and Information Technology, Taif University, Saudi Arabia
Email: marzahrani@tu.edu.sa

Abstract: The duplicate fake accounts are detected in this work the data from the social media platform is accessed. The platform choose to use the analysis on social media platform is selected as twitter. The twitter data is accessed using Twitter API, with using some selected features that remain the most appropriate regarding the reason of duplicate fake account. The feature based analysis is compared using machine learning techniques, Random Forest, Decision Tree, and SVM. The performance is further analyzed based on accuracy SVM performed 93.3% accuracy, where decision tree performed as 89.0% and random forest performed as 85.5%. The better performance observed using feature-based analysis is of SVM.

Keywords: **Feature Based Analysis, Duplicate Account, Account on Twitter, Twitter Duplicate Account**

1. Introduction

The last decade brought out various aspects to get resolve. Using social-media platform is now in trend and every person use social-media platforms. The traffic on the social-media platform is increased exponentially that lead into various necessary issues. The usage of social-media developed various issues in such issues the one of them is developing a duplicate account [1]. The duplicate accounts of the celebrities, high officials, influencers, are creating spam environment on the social media platform. Analyze the duplicate accounts on social-media platforms, especially on the twitter using feature based analysis.

The appropriate features are selected on the basis of the reason that is to detect the duplicate fake accounts. Different techniques are used after selecting the features, like Decision tree, random forest and SVM. However, SVM performed better than the other machine learning techniques.

2. Literature Review:

The ultimate need of the usage of the social media is increase as compared with the last decade. Different social media platforms gained the user traffic exponentially leading towards diverse intentions technically. The user on the social media platforms has diverse culture, locality and user behavior. The accounts of the social media users are being compromised is the one reason and a big question [2]. Similar accounts of the social media platforms are developed by some unknown. Duplicate fake accounts are developed and described as actual accounts even the actual user were neither aware of the duplicate account.

However, platform provides two-way authentications as well but that always take time to get back. There are various such similar concerns within the social media platforms required to address properly. The last decade have introduced various social media platforms like Facebook, Twitter, WhatsApp, WeChat, Telegram, Snapchat and the old ones are the yahoo messenger, outlook, and MSN messenger. These well-known platforms increased diversity on social-media as well as in the routine life [3]. The trends got changed to meeting personally turned to meeting online. This aspect arise the issues to the technical pool as well, to enhance such devices that can manipulate and capable enough to hold such memories.

The need of more memory is arise two decade before and handled well, the next big thing is to handle the issues developed with in such applications launched by developers allowing human interaction to communicate through the devices. The issue like handling such a large scale users is one of them, remaining the interaction integrated is the other one. Furthermore, the globally similar accounts on the social media platforms are another issue on social impacts. Social media platforms required integration like a big name Facebook facing issues between its different products integration [4].

Social media platforms developed such a system to accommodate huge traffic, this remain all possible due to advancement technology made. The usage of social-media platforms increased the technical enhancements, meanwhile the machine learning and artificial intelligence also playing vital role to get major tasks done. Artificial intelligence and machine learning is developing solution from last decade and solving hard problems than ever, this increased in use of artificial intelligence based solutions. Machine learning is evolving diverse nature tasks due to its abilities to handle complex task [5]. Decision based projects involves machine learning technique to consider clear and better results. The

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

social media platforms are evolving through using machine learning techniques.

Artificial intelligence and machine learning both are contributing the social media platforms to enhance their abilities. The analysis on the twitter based on different features contain different reasons are conducted using machine learning techniques. Different aspects lead the attention of the research projects towards the twitter, due to extensive number of users, big data evolution and the approaches used for the data mining projects [6]. In social media platforms twitter itself provide access for the twitter based analysis, leading towards successful contributions. The social media platform provide an environment to generate content in meanwhile, social media influencers are the one who generate content on the social media platforms. Social-media influencers face critical issues related to the similar content published privacy concerns and related other issues that lead to avoid the usage of the social media platform and to create there on private profile [7].

Influential People:

The well-known people usually due to their profession their abilities either by aspect such as designation, including government officials and actors. Their popularity might involve there interest, hobby, either there family background. Such people impact socially on all mediums and develop a clear shadow to the social life. The influencer can be on national level or international level. The most famous examples are the actors, singers, record holders, athletes, high government officials and etc [8]. These people truly represent the community, they are progressive people, they produce leads and people like to follow them or inspires from them. Influential people are mostly the trend setters, or involved in setting trends.

Social Media Influencers:

The social media influencers can be differentiate by different means and classified in different categories somehow to maintain the progress here are some of the social media influencers mentioned and explained well. The different types of social media influencers are social activists, bloggers, photographers, micro-influencers, nano-influencers, journalists, and celebrities [9].

Types of Social Media Influencers	**Social activists:** The people communicate using only social media platforms and having an active social media account.
Bloggers: The people having a dedicated blog publish or the related content as of the niche of the blog.	**Photographers:** These accounts holders just publish the photographs and there photos speaks with influence**.**
Micro- Influencers: The influencers with millions of the followers, on the social media platforms are known as the micro influencers.	**Nano- Influencers:** The influencers with few thousands of followers on social media platforms are known as nano-influencers.
Journalists: The journalists on social media mostly sets the trends by explaining the new trends, they are a very good influencers.	**Celebrities:** The well-known people are also influencers as the people like to follow them, and these may be totally from different fields or areas.

Impact on life of social-media platforms:

The social-media platforms impacts on the entire life leading to remain update with the global village. Every person is connected through social media to communicate or to remain updated with the entire world. Ethically or unethically both ways social media impacts on the life the surveys explains that 80% people are addicted to the social media platforms and uses social media platforms on daily basis [10]. Utilizing the free time on the social-media platforms and exploring into different aspect lead into increased usage of social media platforms.

Social-media Influencers Impact on Social-media:

The influencers are the trend setters on the social media platforms, sometimes they lead a challenge to influence, and sometimes they explain the new happening around there geographical place. Similarly, there are influencers involved even making routine things some of them express their thoughts using pictures they take, some influencer’s impact by saying about a specific issue within the society, some influencers influence by giving decision about the happening or a new thing. So these all types of the influencers impact on the society with expressing their thoughts. Similarly, these influencer’s impacts on the social media as well as on the life’s of the single individuals with their different perspective [11]. Different types of the social media influencers impact on the social media by different perspective but they impact on the social media platforms and to the lives of the social media users.

Issue Faced by Social-media influencers:

The issue meanwhile these social media influencers face is the duplicate account they never created. Meanwhile, the influencers are busy in creating the content, regarding their specific area, they were never aware of having any duplicate account having the same name same content and totally spamming with the related people. This lead into real time issues for the social media influencers. The duplicate account is further used for the different purpose some of the fake duplicate accounts lead to generate some fake identity and then manipulate fake content or get involved in fake discussions within chat [12]. These are all set of issues involved with single reasons of generating or making the fake duplicate account on the social-media platforms of the social media influencers.

Duplicate Fake Accounts of Influencers on Social-media Platforms:

The duplicate fake accounts of influencers are creating a gap to be considered to detect them and to eliminate them. The creators of the duplicate accounts are not familiar with the issues created after creating the duplicate accounts. Similarly, using machine learning and artificial intelligence based solutions this can be detected. Social-media platforms are still unable to provide a clear platform free from issues. The mega influencers face duplicate fake account issue that directly impose issues diversely [13]. Social-media platforms are dedicatedly providing a platform for communication across the oceans somehow some issues require a proper attention. Duplicate fake accounts on social media platforms are one of the serious issues and require to be considered properly.

Twitter Based Duplicate fake accounts:

The social media platforms have duplicate fake accounts developing an area to be considered in resolving the issue. On twitter duplicate fake accounts of the users are created and manipulate diverse intentions. Even the creators were neither aware of the duplicate account, that is used with different intentions. Twitters overcome this issue after the actual account holder or actual user complains, this may take a span of time to eliminate the issue [14]. There are still two factor impact this issue first one is that one who is unaware of having any duplicate account. Second there is still a time require to lunch complain to restrict that fake duplicate user. The duplicate fake account issue arises with the influential people based on twitter.

Different Approaches to Detect Duplicate Fake Account:

The very basic technique to check is manually however while in the technical era, the twitter based fake account can be analyzed using machine learning techniques. The one of them is detection of duplicate fake accounts is using feature based detection, the other way is to detect it using keywords used to search and the third way to detect is using the tool based search.

However, the different purpose of the fake account detection matters and the most exclusive way to detect the fake duplicate accounts is to detect using the feature based analysis. The other methods were in use previously, but the recent method used for the twitter based duplicate fake accounts is feature based analysis.

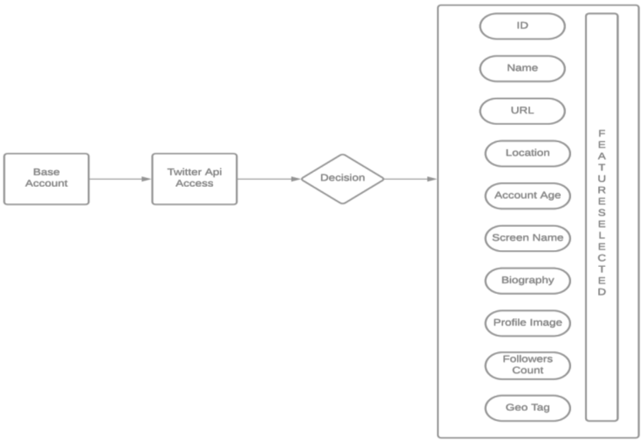
Proposed Method Used to detect the duplicate account on twitter:

A comparison based analysis is conducted to detect the duplicate account on twitter used for the different purpose. Mainly, the specific features are used to detect the duplicate accounts, this lead twitter is accessed using the twitter API on the basis of the features selected then the results are generated on comparison of the base account with the twitter data. Different methods are already working progressively in this work three of the machine learning techniques are used to perform better [15]. Support vector machine is used to detect the duplicate fake accounts, decision tree and the random forest is used to detect the duplicate fake account on the twitter using API.

Twitter Api Access:

The twitter API is accessed using python language using developer.twitter.com to access twitter data. Step by step process for the access of twitter API is to create account on the developer site and then to request the access for research based task will lead to the access to the twitter data.

The below image explains how the twitter API is accessed for this task, firstly a base account is created, then using twitter API is accessed the specific twitter features and the related data is fetched [16].



Twitter API Access

Features selected:

The features used for the progressive analysis using machine learning techniques, are selected specifically for the duplicate account detection on the twitter. These features are selected with the reason involved in the detection of the fake accounts these are the most related features. The selected features are ID, Name, URL, Location, Age of the user account, Screen Name, Biography, Profile Image, Followers Count, and Geo Location.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Features Selected	Description
ID	Unique ID given to every twitter account holder.
Name	Named used in profile
URL	URL used in profile.
Location	Location of the account.
Age of the user Account	Age when account started from.
Screen Name	Plain English name.
Biography	Accounts biography section.
Profile image	Profile image used.
Followers Count	Followers account holds.
Geo tag	Tags of the locations account used.

Machine Learning Algorithms:

The comparison is based on the machine learning techniques used in the feature based models. Twitter is accessed using twitter API initially an account is used to find the duplicate fake account, the criteria remains the same for all three classifiers [17].

Random Forest:

The random forest use this equation while progressively used for the duplicate fake account detection on the twitter. The twitter data is accessed using twitter API, random forest performed better to detect the fake duplicate accounts [18].

$$\boldsymbol{MS = \frac{1}{N} \sum_{I=1}^{N} (Fi - Yi)2}$$

MS stands for the mean square of the features, N is the number of features used to detect the duplicate account, F is the progressive value of returned from the model, Y is the data record searched for.

Support Vector Machine:

The detection of the duplicate fake account on twitter using features based analysis, SVM is also used. SVM basically divide using the vector, helps in developing the decisions using hyper planes [19]. It explains the close data points from the hyperplane, the information retrieval like data points against the features selected for duplicate fake accounts.

$$\mathrm{H: w \bullet xi + b = 1}$$

In these equations H are the supports however the phenomena of the machine learning technique W explains about the weight, X is the input vector, and B explains about the biasness of the decision developed.

Decision Tree:

The different variants of the decision tree are used to develop the appropriate solutions using decision tree entropy equation is used to analyze the twitter based analysis. The duplicate fake accounts are observed using the Decision tree.

$$E(S) = \sum_{I=1}^{C} -Pi\ Log2\ Pi$$

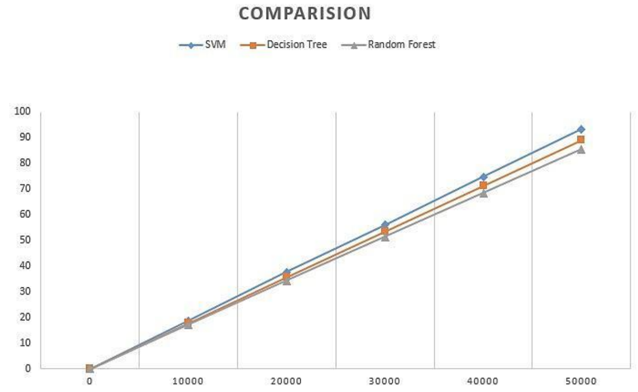
In the above equation, Pi represents the probability of having duplicate fake account, where S is the current feature, and I represent the node in the feature S [20].

Comparison Results Based on Accuracy:

The implementation of the three machine learning based algorithms use to detect the twitter based duplicate accounts. The comparison is conducted using twitter based analysis, based on the selected features

Techniques	SVM	Decision Tree	Random Forest
Accuracy	93.3%	89.0%	85.5%
Account Data Limit	50,000	50,000	50,000

The observed accuracy is on the set of the 50,000 accounts on all next iteration that lead an appropriate search against the duplicate fake account. Every set of iteration is considered with maximum figure of 50,000 accounts. In this situation SVM performed better as 93.3% and decision tree came out with 89.0% and random forest came out with 85.5%.



Acknowledge:

The programming support was provided by Muhammad Ehtisham moreover, the language used to code in python and R. The author wants to thank you for the contribution of their support in implementation.

Conclusion:

To detect the duplicate fake accounts, different features are used the features selected on the importance based on most relevant regarding the issue of duplicate fake accounts. Meanwhile, the selected features created ease in detecting the fake accounts, SVM produce impressive results instead of Random forest and decision tree. The improved results are 93.3% using SVM, and using the specific features selected on the basis of the importance involved in duplicate fake accounts on twitter.

References

[1] Herzallah, W., Faris, H., & Adwan, O. (2018). Feature engineering for detecting spammers on twitter: Modelling and analysis. Journal of Information Science, 44(2), 230-247.

[2] Erşahin, B., Aktaş, Ö., Kılınç, D., & Akyol, C. (2017, October). Twitter fake account detection. In 2017 International Conference on Computer Science and Engineering (UBMK) (pp. 388-392). IEEE.

[3] Madisetty, S., & Desarkar, M. S. (2018). A neural network-based ensemble approach for spam detection in Twitter. IEEE Transactions on Computational Social Systems, 5(4), 973-984.

[4] Loyola-González, O., Monroy, R., Rodríguez, J., López-Cuevas, A., & Mata-Sánchez, J. I. (2019). Contrast pattern-based classification for bot detection on twitter. IEEE Access, 7, 45800-45817.

[5] Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2015). Fame for sale: Efficient detection of fake Twitter followers. Decision Support Systems, 80, 56-71.

[6] Boichak, O., Hemsley, J., Jackson, S., Tromble, R., & Tanupabrungsun, S. (2021). Not the Bots You Are Looking For: Patterns and Effects of Orchestrated Interventions in the US and German Elections. International Journal of Communication, 15, 26.

[7] Kaur, A., & Sinha, A. (2021). Multi-contextual spammer detection for online social networks. Journal of Discrete Mathematical Sciences and Cryptography, 1-10.

[8] Permana, F. C., Wicaksono, Z. M., Kurniawan, C., Abdullah, A. S., & Ruchjana, B. N. (2021, January). Perception analysis of the Indonesian society on twitter social media on the increase in BPJS kesehatan contribution in the Covid 19 pandemic era. In Journal of Physics: Conference Series (Vol. 1722, No. 1, p. 012022). IOP Publishing.

[9] Atodiresei, C. S., Tănăselea, A., & Iftene, A. (2018). Identifying fake news and fake users on Twitter. Procedia Computer Science, 126, 451-461.

[10] Fitriani, Y., Sumpeno, S., & Purnomo, M. H. (2019, April). Classifying Twitter Spammer based on User's Behavior using Decision Tree. In 2019 Asia Pacific Conference on Research in Industrial and Systems Engineering (APCoRISE) (pp. 1-7). IEEE.

[11] Khalil, A., Hajjdiab, H., & Al-Qirim, N. (2017). Detecting fake followers in twitter: A machine learning approach. International Journal of Machine Learning and Computing, 7(6), 198-202.

[12] VanDam, C., Tang, J., & Tan, P. N. (2017, August). Understanding compromised accounts on twitter. In Proceedings of the International Conference on Web Intelligence (pp. 737-744).

[13] Volkova, S., & Bell, E. (2017, May). Identifying effective signals to predict deleted and suspended accounts on twitter across languages. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 11, No. 1).

[14] Erşahin, B., Aktaş, Ö., Kılınç, D., & Akyol, C. (2017, October). Twitter fake account detection. In *2017 International Conference on Computer Science and Engineering (UBMK)* (pp. 388-392). IEEE.

[15] Das, S., Behera, R. K., & Rath, S. K. (2018). Real-time sentiment analysis of twitter streaming data for stock prediction. *Procedia computer science*, *132*, 956-964.

[16] Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance*, *6*(2), e19273.

[17] Trupthi, M., Pabboju, S., & Narasimha, G. (2017, January). Sentiment analysis on twitter using streaming API. In *2017 IEEE 7th International Advance Computing Conference (IACC)* (pp. 915-919). IEEE.

[18] Munjal, P., Narula, M., Kumar, S., & Banati, H. (2018). Twitter sentiments based suggestive framework to predict trends. *Journal of Statistics and Management Systems*, *21*(4), 685-693.

[19] Campan, A., Atnafu, T., Truta, T. M., & Nolan, J. (2018, December). Is data collection through twitter streaming api useful for academic research?. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 3638-3643). IEEE.

[20] Nagdeve, A. S., & Ambekar, M. M. (2020, October). Spam Detection by designing Machine Learning approach in Twitter Stream. In *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)* (pp. 126-130). IEEE.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731