

# Fake Instagram Profile Detection and Classification Using Machine Learning

<sup>1</sup>Mrs.B. SHIVANI, <sup>2</sup>T. TEJA SRI, <sup>3</sup>N. NIKHITHA, <sup>4</sup>M.A. AHAD SIDDIQUI

<sup>1</sup>Assistant Professor, Dept.of IT, TKR College of Engineering & Technology, Meerpet, Hyderabad,  
[bhutam.shivani@gmail.com](mailto:bhutam.shivani@gmail.com)

<sup>2</sup>BTech student, Dept.of IT, TKR College of Engineering & Technology, Meerpet, Hyderabad,  
[tejasrithurupu@gmail.com](mailto:tejasrithurupu@gmail.com)

<sup>3</sup>BTech student, Dept.of IT, TKR College of Engineering & Technology, Meerpet, Hyderabad,  
[nnikhitha900@gmail.com](mailto:nnikhitha900@gmail.com)

<sup>4</sup>BTech student, Dept.of IT, TKR College of Engineering & Technology, Meerpet, Hyderabad,  
[ahadsiddiqui042@gmail.com](mailto:ahadsiddiqui042@gmail.com)

**Abstract:** *With the increase in Internet usage, Instagram is now considered a very important platform for advertising marketing and social interaction. It is used by millions of users but, some users tend to misuse the platform by creating false identities. Moreover, the popularity of social media users is determined by followers and hence users resort to different wrong means to promote increased profile followers. Fake engagement is one of the significant problems in Online Social Networks (OSNs) which is used to increase the popularity of an account in an inorganic manner. The detection of fake engagement is crucial because it leads to loss of money for businesses, wrong audience targeting in advertising, wrong product predictions systems, and unhealthy social network environment. This study is related with the detection of fake and automated accounts which leads to fake engagement on Instagram. Prior to this work, there were no publicly available dataset for fake and automated accounts. For the detection of these accounts, machine learning algorithms like Naive Bayes, Logistic Regression, Support Vector Machines and Neural Networks are applied. Additionally, for the detection of automated accounts, cost sensitive genetic algorithm is proposed to handle the unnatural bias in the dataset. To deal with the unevenness problem in the fake dataset, Smote-nc algorithm is implemented. For the automated and fake account detection datasets, 86% and 96% classification accuracies are obtained, respectively.*

**Keywords:** *Online Social Networks, Machine learning, Instagram, fake profile identification.*

## I. INTRODUCTION

With the arrival of the Internet and social media, at the same time as masses of humans have benefitted from the full-size re assets of records available, there was a full-size boom with inside the upward push of cyber-crimes, mainly targeted closer to women. According to a 2019 file with inside the Economics Times, India has witnessed a 457% upward push in cybercrime with inside the 5 years span among 2011 and 2016. Most speculate that that is because of effect of social media inclusive of Facebook, Instagram and Twitter on our day by day lives. While those simply assist in growing a legitimate social network, advent of consumer debts in those websites normally desires simply an email-id. A actual lifestyles man or woman can create more than one fake IDs and for this reason impostors can effortlessly be made. Unlike the actual international state of affairs in which more than one policies and guidelines are imposed to become aware of oneself in a completely unique manner (as an instance at the same time as issuing one's passport or driver's license), with inside the digital international of social media, admission does now no longer require this kind of checks. In this paper, we study the one-of-a-kind debts of Instagram, specifically and try and verify an account as fake or actual the use of Machine Learning strategies

specifically Logistic Regression and Random Forest Algorithm.

Instagram is an online photo and video sharing social networking platform that has been available on both Android and iOS since 2012. As of May 2019, there are over a billion users registered on Instagram. In the recent years, Instagram has been found to be using third party apps, called bots. While these can definitely impersonate a user and tarnish their reputation leading to 'identity theft', there has also been greater instances of malicious ways of promoting the brand image of a company known as "influencer marketing". These days a number of businesses are using social media to heed to their customers' needs which has led to yet another malpractice called Angler phishing. All these malpractices have made it vital to implement strong fraud detection techniques and hence we propose our solution.

## II. LITERATURE SURVEY

Today, social networks are developing at an incredible speed. These services are vital to the human masses in society, especially to advertising campaigns, celebrities, and politicians trying to market themselves using fans and fanatics on social media. Therefore, fake money owed

created in the name of human beings and corporations can be dangerous, reputational damage to human and organization, and in the long run hastened the downgrading of their true likes and fans. Moreover, all kinds of fake profiles have a detrimental effect on the social media benefits of advertising, marketing and advertising organizations. These fake profiles can be a form of cyberbullying; Real customers also have great concerns about their privacy in the online environment with those fake profiles.

Therefore, in recent years, many researchers have investigated the problem of detecting malicious sports and spammers on social media using system scanning strategies. However, there is a limited pattern of research articles on detecting phantom debt or fake fanatics. In this section, we highlight every spammer and fake debt answer that has been added these days.

Ferrara et al. It provided a way to hit bot users on Twitter based entirely on the somewhat common powers that set them apart from valid users. In their proposed approach, they used a system that defines the method and behavior patterns between valid debt and bot debt to classify the money owed on the bot or valid bill.

Cresci et al. have created and used a baseline dataset of verified human and fake fans on Twitter. In their work, they exploited the baseline dataset to educate a fixed of gadget gaining knowledge of classifiers constructed primarily based totally on reviewed policies and capabilities set the use of the media. Their proposed approach is green in detecting fake money owed; the consequences accomplished via way of means of their approach display it can classify greater than 95% of the money owed successfully from the authentic schooling set.

In a barely extraordinary approach, Zhang and Lu. Introduced a unique approach for the detection of fake money owed in Weibo. Their proposed answer has extraordinary aspects. At first, that they'd this premises why such money owed exist with inside the first place. In the second, they investigated the overlap among fans listing of the clients of fake fans, and that they located an excessive overlap among their follower lists. Their research located 395 nearduplicates, which caused 11.90 million fake monies owed that despatched 1,000,000 hyperlinks with inside the network.

Thomas et al. made a group of 1.8 million tweets despatched via way of means of 32.9 Twitter money owed. In their research,

they located Twitter suspended approximately 1.1 million of these money owed. They have decided on randomly a hundred of these money owed to research their tweets and confirm they had been spamming money owed.

They made a similarly evaluation on that a hundred decided on money owed, and that they locate 93 of the chosen money owed had been suspended for posting junk mail and the unsolicited commercial of numerous products. Three different monies owed had been suspended for re-tweeting content material of extraordinary information money owed, and the alternative four remained money owed had been suspended for reproduction and competitive advertising posts.

Gao et al. have used a fixed of capabilities for efficiently reconstructing junk mail tweets into campaigns in place of studying them separately. The end result suggests their proposed answer received. However, the disadvantage in their approach is its low detection accuracy.

Benevenuto et al. proposed a option to stumble on spammers from non spammers. In their approach, they used an SVM classifier, that is a supervised gadget gaining knowledge of algorithm. They have used 23 conduct and 39 content

material capabilities to distinguish spammers from nonspammers, and that they accomplished experiments via way of means of 5-fold cross-validations. The experiments display they had been nearly a success in figuring out spammers from non-spammers.

Bala Anand et al. advanced a new gadget to stumble on fake customers at the Twitter platform the use of a graph-primarily based totally semi-supervised gaining knowledge of algorithm (EGSLA) and examine and amassing behavioral and person-generated content material (UGC) information. The version first gathered customers information, analyzed them to extract beneficial capabilities, and then accomplished type on those capabilities and made decisions. The experimental consequences display that the EGSLA algorithm accomplished excessive overall performance and turned into greater useful than different algorithms inclusive of choice tree, KNN, SVM, and game theory-primarily based totally techniques in phrases of type accuracy.

**Sahoo et al.** Provided a hybrid model for finding malicious social media profiles by targeting Twitter. The proposed hybrid version consists of modules; First, they analyzed and extracted the efficiencies with the internet form of Petrie, and then

used the capabilities of those efficiencies as the classifiers come to categorize the profiles as malicious and legitimate instructions. Experimental results show that the proposed method effectively outperformed Twitter's money owed and obtained a high detection rate in type-accuracy words. Therefore, according to the literature, many researchers have used the devices to gain knowledge about techniques to overcome security problems in social networks. The research surveyed generally focused on spotting spam on microblog social networks. They have researched many solutions to solve the problem of spam and fake money owed on Twitter and the amazing microblogging social network. However, as of now, there is no complete solution to phantom debt on the Instagram platform, and this is one of the motivations at the end of this study. Therefore, in this paper, we propose an inexperienced approach to detect fake debts on the Instagram platform, which can correctly classify the large money owed by Instagram people.

### III. PROPOSED WORK

In this paper, the automatic detection of fake profiles has been proposed to identify fake Instagram profiles so that the social life of Instagram users is secure. The prediction of fake Instagram profiles is

facilitated using supervised learning machine algorithms. Upon classification, fake profile IDs are stored in a data dictionary to further help the concerned authorities to take necessary actions against fraudulent social media profiles. Experimentation has been done to compare the classification algorithms used to train the dataset. The factors used by the existing systems to detect the fake accounts are very less. The prediction becomes accurate when the number of parameters used are more efficient. In previously used algorithms, if some of the inputs are not appropriate, the algorithm could not produce accurate results. Hence, in this research, we made use of gradient boosting algorithm. It uses decision trees as a prime factor. We made use of several parameters. These parameters are further considered as inputs which are used to form the decision trees in order to apply gradient boosting algorithm.

### SYSTEM ARCHITECTURE

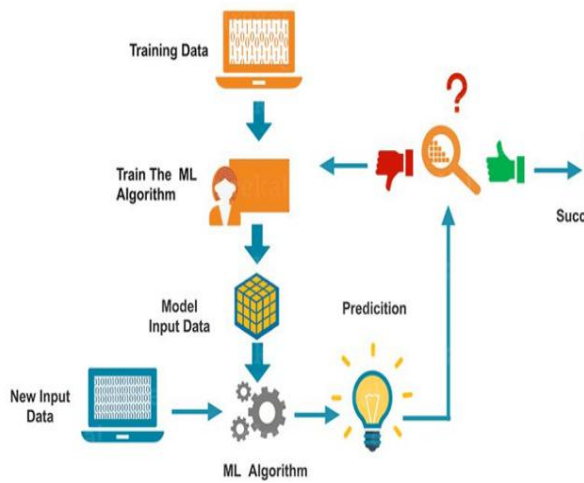


Fig.1 System architecture

### GATHERING DATA

Data Gathering is the first step of the machine learning life cycle. The goal of this step is to identify and obtain all data-related problems.

In this step, we need to identify the different data sources, as data can be collected from Kaggle such as csv files. It is one of the most important steps of the life cycle. The quantity and quality of the collected data will determine the efficiency of the output. The more will be the data, the more accurate will be the prediction.

This step includes the below tasks:

- Identify various data sources
- Collect data
- Integrate the data obtained from different sources

By performing the above task, we get a coherent set of data, also called as a dataset. It will be used in further steps.

### DATA PREPARATION

After collecting the data, we need to prepare it for further steps. Data preparation is a step where we put our data into a suitable place and prepare it to use in our machine learning training.

In this step, first, we put all data together, and then randomize the ordering of data.

This step can be further divided into two processes:

- Data exploration:

It is used to understand the nature of data that we have to work with. We need to understand the characteristics, format, and quality of data.

A better understanding of data leads to an effective outcome. In this, we find Correlations, general trends, and outliers.

- Data pre-processing:

Now the next step is pre-processing of data for its analysis.

### DATA WRANGLING

Data wrangling is the process of cleaning and converting raw data into a useable format. It is the process of cleaning the

data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step. It is one of the most important steps of the complete process. Cleaning of data is required to address the quality issues.

It is not necessary that data we have collected is always of our use as some of the data may not be useful. In real-world applications, collected data may have various issues, including:

- Missing Values
- Duplicate data
- Invalid data
- Noise

So, we use various filtering techniques to clean the data.

It is mandatory to detect and remove the above issues because it can negatively affect the quality of the outcome.

## **DATA ANALYSIS**

Now the cleaned and prepared data is passed on to the analysis step. This step involves:

- Selection of analytical techniques
- Building models
- Review the result

The aim of this step is to build a machine learning model to analyze the data using various analytical techniques and review the outcome. It starts with the determination of the type of the problems, where we select the machine learning techniques such as Classification. then build the model using prepared data, and evaluate the model.

Hence, in this step, we take the data and use machine learning algorithms to build the model.

## **TRAIN MODEL**

Now the next step is to train the model, in this step we train our model to improve its performance for better outcome of the problem.

We use datasets to train the model using various machine learning algorithms. Training a model is required so that it can understand the various patterns, rules, and, features.

## **TEST MODEL**

Once our machine learning model has been trained on a given dataset, then we test the model. In this step, we check for the accuracy of our model by providing a test dataset to it. Testing the model determines the percentage accuracy of the model as per the requirement of project or problem.



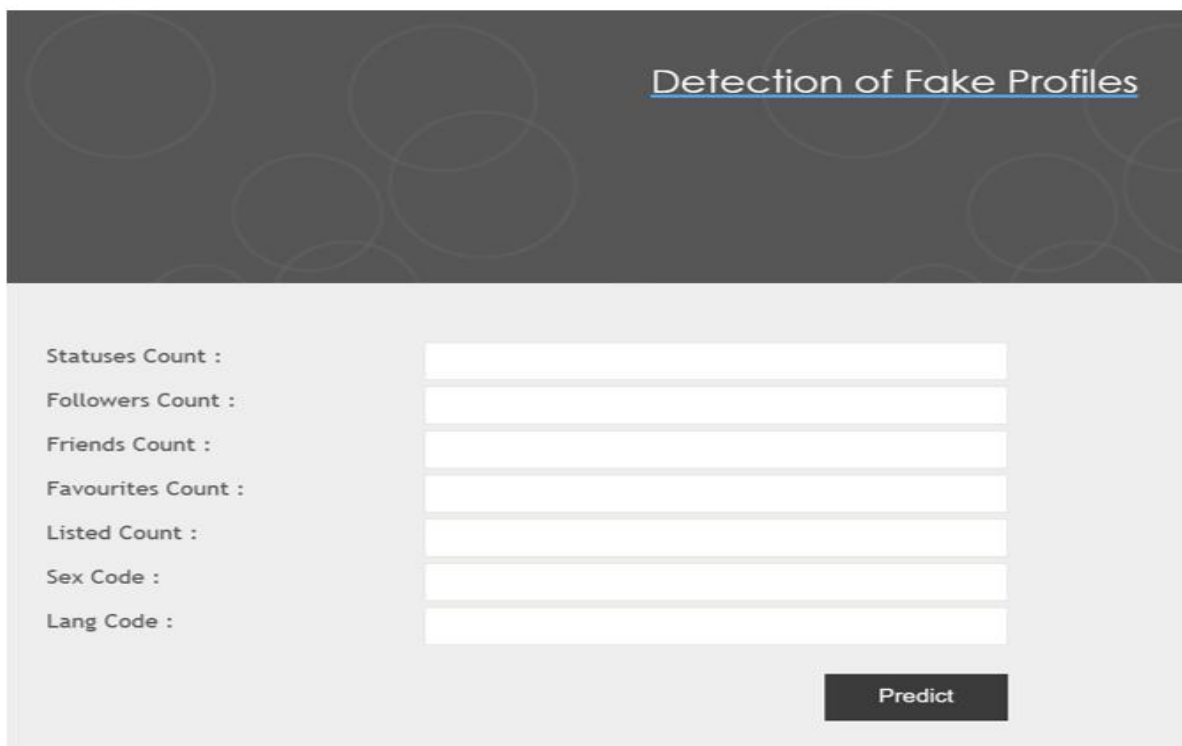
## DEPLOYMENT

The last step of machine learning life cycle is deployment, where we deploy the model in the real-world system.

If the above-prepared model is producing an accurate result as per our requirement

with acceptable speed, then we deploy the model in the real system. But before deploying the project, we will check whether it is improving its performance using available data or not. The deployment phase is similar to making the final report for a project.

## IV. RESULTS



Detection of Fake Profiles

Statuses Count :

Followers Count :

Friends Count :

Favourites Count :

Listed Count :

Sex Code :

Lang Code :

Predict

Fig.2 Give the input to detect the profile is fake or real



Detection of Fake Profiles

Statuses Count :	<input type="text" value="20"/>
Followers Count :	<input type="text" value="3500"/>
Friends Count :	<input type="text" value="2000"/>
Favourites Count :	<input type="text" value="250"/>
Listed Count :	<input type="text" value="28"/>
Sex Code :	<input type="text" value="1"/>
Lang Code :	<input type="text" value="1"/>

Fig.3 Given inputs to detect the profile is fake or real

Detection of Fake Profiles

**Real User**

Statuses Count :	<input type="text"/>
Followers Count :	<input type="text"/>
Friends Count :	<input type="text"/>
Favourites Count :	<input type="text"/>
Listed Count :	<input type="text"/>
Sex Code :	<input type="text"/>
Lang Code :	<input type="text"/>

Fig.4 Real User

Detection of Fake Profiles

Statuses Count :	<input type="text" value="20"/>
Followers Count :	<input type="text" value="3500"/>
Friends Count :	<input type="text" value="1000"/>
Favourites Count :	<input type="text" value="0"/>
Listed Count :	<input type="text" value="80"/>
Sex Code :	<input type="text" value="1"/>
Lang Code :	<input type="text" value="1"/>

Fig.5 Given inputs to detect the profile is fake or real

Detection of Fake Profiles

**Fake User**

Statuses Count :	<input type="text"/>
Followers Count :	<input type="text"/>
Friends Count :	<input type="text"/>
Favourites Count :	<input type="text"/>
Listed Count :	<input type="text"/>
Sex Code :	<input type="text"/>
Lang Code :	<input type="text"/>

Fig.6 Fake User

## V. CONCLUSION

A new classification algorithm was proposed to improve detecting fake accounts on social networks, where the RANDOM FOREST trained model decision values were used to train a Neural Network model. To reach our goal we used dataset run it into the pre-processing phase where different feature reduction techniques were used to reduce the feature vector. In the classification phase, random forest learning algorithms were used. The results of the analyses showed that random forest has archived better accuracy results with all feature sets comparing with other classifiers, with classification accuracy around 98%. It was noticed that the Neural Network algorithm has the lowest classification accuracy compared with random forest.

In case some of the inputs are missing, it gets so tedious for random forest algorithm to give the results. So, gradient boosting algorithm is used to detect the account is fake or real even some inputs are missing. The Accuracy of detecting fake accounts is found to be higher using Gradient Boosting Algorithm followed by Random Forest Algorithm for a given dataset. So,

here we are using gradient boosting algorithm to detect accurate output.

## REFERENCES

- [1] Ramalingam, D., Chinnaiah, V. (2018). Fake profile detection techniques in large scale online social networks.
- [2] Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7): 96-104.
- [3] Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M. (2015). Fame for sale: Efficient detection of fake Twitter followers. *Decision Support Systems*.
- [4] Zhang, Y., Lu, J. (2016). Discover millions of fake followers in Weibo. *Social Network Analysis and Mining*.
- [5] Thomas, K., Grier, C., Song, D., Paxson, V. (2011). Suspended accounts in retrospect: An analysis of twitter spam. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, pp.243-258.
- [6] Benevenuto, F., Magno, G., Rodrigues, T., Almeida, V. (2010). Detecting spammers on twitter. In *Collaboration, Electronic Messaging, Anti-abuse and Spam Conference (CEAS)*, p. 12

- [7] Schoonjans, F. (2019). ROC curve analysis with MedCalc. [Online] MedCalc. Available at: <https://www.medcalc.org/manual/roc-curves.php> [Accessed 10 Jun. 2019].
- [8] Kietzmann, J.H., Hermkens, K., McCarthy, I.P., Silvestre, B.S., 2011. Social media? Get serious! Understanding the functional building blocks of social media. *Bus.Horiz., SPECIAL ISSUE: SOCIAL MEDIA* 54, 241251. doi: 10.1016/j.bushor.2011.01.005.
- [9] Krombholz, K., Hobel, H., Huber, M., Weippl, E., 2015. Advanced Social Engineering Attacks. *J Inf SecurAppl* 22, 113–122. doi: 10.1016/j.jisa.2014.09.005