ISSN: 2321-2152

# IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

# FAKE ACCOUNT DETECTION USING DS & ML

[1] DR. AMITHKUMAR, [2] R. SUMA, [3] CH. SRIKANTH, [4] K. GEETHA, [5] SYED NAWAZ

[2,3,4,5] U.G. Scholor, Department of DS, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.

[1]Associate Professor, Department of DS, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.

***Abstract-*** Nowadays, Online Social Media is dominating the world in several ways. Day by day the number of users using social media is increasing drastically. The main advantage of online social media is that we can connect to people easily and communicatewith them in a betterway. Thisprovided a newway of a potential attack, such as fake identity, false information, etc. A recent survey suggest that the number of accounts present in the social media is much greater than the users using it. This suggest that fake accounts have been increased in the recent years. Online social media providers face difficulty in identifying these fake accounts. The need for identifying these fake accounts is that social media is flooded with false information, advertisements, etc.

Traditional methods cannot distinguish between real and fake accounts efficiently. Improvement in fake account creation made the previous works outdated. The new models created used different approaches such as automatic posts or comments, spreading false information or spam with advertisements to identify fake accounts. Due to the increase in the creation of the fake accounts different algorithms with different attributes are use. Previously use algorithms like naïve bayes, support vector machine, random forest has become inefficient in finding thefake accounts. In this research, we came up with an innovative method to identify fake accounts. We used gradient boosting algorithm with decision tree containing three attributes. Those attributes are spam commenting, artificial activity and engagement rate. We combined Machine learning and Data Science to accurately predict fake accounts.

***Keyword:Data science, Fake account detection, Machine learning, Online social media***

## I. INTRODUCTION

In today's Modern society, social media plays a vital role in everyone's life. The general purpose of social media is to keepintouchwithfriends,sharingnews,etc.Thenumberof users in social media is increasing exponentially. Instagram hasrecentlygained immense popularityamong social media users. With more than 1 Billion active users, Instagram has become one of the most used social media sites. After the emergence ofInstagramto the social media scenario, people with a good number of followers have been called Social MediaInfluencers.Thesesocialmediainfluencershavenow become a go-to place for the business organization to advertise their products and services.

RevisedManuscriptReceivedonNovember30,2019.
*CorrespondenceAuthor
**S.P.Maniraj***,AssistantProfessor,SRMInstituteofScienceandTechnology, Chennai, India.
**HarieKrishnanG**,UGScholar,SRMInstituteofScienceandTechnology, Chennai, India.
**SuryaT**,UGScholarSRMInstituteofScienceandTechnology,Chennai, India.
**PranavR**,UGScholarSRMInstituteofScienceandTechnology,Chennai, India.



The widespread useofsocial mediahasbecomebothaboon and a bane for the society. Using Social media for online fraud, spreading False information is increasing at a rapid pace. Fake accounts are the major source of falseinformation on social media. Business organizations that invest huge Sumof moneyon social media influencers must know whether the following gained by that account is organic or not. So, there is a widespread need for a fake accountdetectiontool,whichcanaccuratelysay whetherthe account is fake or not.In this paper, we use classification algorithms in machine learning to detect fake accounts. The process of finding a fake account mainly depends on factors such as engagement rate and artificial activity.


**Fig1.1GraphShowingincreaseinnumberofFake accounts over the years**

## II. EXISTINGSYSTEM

The existing systems use very fewer factors to decide whether an account is fake or not. The factors largely affect the way decision making occurs. When the number offactors is low, the accuracy of the decision making is reduced significantly.There is an exceptional improvement in fake account creation, which is unmatched by thesoftware or application used to detect the fake account. Due to the advancement in creation of fake account, existing methods have turned obsolete. The most common algorithm used by fake account detection Applications is the Random forest algorithm. The algorithm has few downsides such as inefficiency to handle the categorical variables which has different number of levels. Also, when there is an increasein the number of trees, the algorithm's time efficiency takesa hit.

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

## III. PROPOSEDSYSTEM

Theexistingsystemusesrandomforestalgorithmtoidentify the fakeaccount.It isefficient whenit has the correctinputs and when it has all the inputs. When some of the inputs are missingit becomes difficult for the algorithmto produce the output. To overcome such difficulties in the proposed systems we used a gradient boosting algorithm. Gradient boosting algorithm is like random forest algorithm which uses decision trees as its main component. We also changed the way we find the fake accounts i.e., we introduced new methods to find the account. The methods used are spam commenting, engagement rate and artificial activity. These inputs are used to form decision trees that are used in the gradient boosting algorithm. This algorithm gives us an output even if some inputs are missing. This is the major reason for choosing this algorithm. Due to the use of this algorithm we were able to get highly accurate results.

## IV. DETECTIONSTRATEGY

In Our Research, we define an account as fake when it doesn't meet the minimum engagement rate, have artificial activities or when the account has a history of Spam comments.

### A. WebScraper

Web Scraper is used to extract data from a website. When a user pastes a link of a Social media Account, Using OutWit hub, a Web scraper tool, we extract necessary pieces of information from the social media site.

We extract data such as login activity, Total Likes, Total Comments, Number of posts, Number of followings and Number of Followers.

### B. CalculationofEngagementrate

An engagement rate is a metric that measures the level of engagementofaPostorStoryreceivedonsocialmedia.Itis the percentage by which the audience interact with a post.By checking the number of interactions with the number of followers we can evaluate the engagement rate. Interactions can be of likes, comments, and shares. Most Fake accounts will boast of 1000s of followers and a very minimum number of likes. Since the engagement rate is relatively calculated, comparisons between popular accounts andsemi-popular accounts are comparatively easy. This metricis one of the most vital ones because lesser audience engagement signifies that the account is fake.

$$\text{Engagement rate percentage} = \left( \frac{\text{Total number of interactions}}{\text{Total number of followers}} \right) \times 100$$

**Fig4.1EngagementrateCalculation**

### C. ArtificialActivity

Normal social media activities such as liking, commenting and sharing turns into an artificial activity when the frequencyoftheabove mentionedareveryhigh. Aroundthe clock activity also signifies that the account is used by aBot. At this stage, we look into the number of likes, comments,andsharesthisparticularaccounthasmadesince itscreation.Ifanenormousamountoflikesorcommentsare found, then that account will be considered as fake. By enormous we mean a number which is not achievable by an average social media user. Also, the amount of time the account was online will be looked upon before concluding. Otherfactorsthatareconsideredareinsufficientinformation on the account and Status of verification of the mobile number and email.

### D. SpamComments

BOT comments are always known to be very Generic and often lack Substance. At this stage, the comments madefromthe account will be gone through in a detailed manner. Total number of comments by the user madesince the creation of the account will be compared with average comments of users in that particular OSN's. If there is a big differencetheaccountmaybeconsideredfake. Commenting links will lead to the account being termed as Fake account. Same or Similar type of comments will also be consideredas spam comments.

### E. DetectionofFakeAccounts

In this step, we combine all the data we extracted from the website. In this paper we mainly focus on engagement rate, artificial activity and spam comments. The data collected using web scraper is used to compute the values for the factors mentioned above. Using these factors different decision tress is formed. Using gradient boosting algorithm and with the formed decision trees fake accounts are detected.

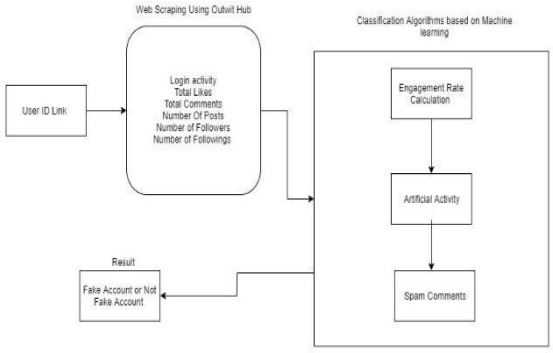


**Fig4.2Architecture Diagram**

## V. EVALUATION

### A. DecisionTrees

Decision trees are made seeing the success rate i.e., in our case taking the value which contains more fake accounts. The first tree is made using an engagement rate as the root node and artificial activityas its child node along withspam commentsasanothernode.Thesecondtreeis made keeping artificial activity as the root node, engagement rate andspam comments as subsequent nodes. The third tree is formed using spam comments as the root node, artificial activity and engagement rate as subsequent nodes.

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

**B. Gradientboosting**

Gradient boosting is the most effective algorithm for classification problems. When the values are given accurately and with a lot of training data sets this algorithm worksefficiently. The basicprinciple ofgradientboostingis that it forms a strong rule from multiple weak learners. The main advantage of this algorithm is that it predicts perfectly in the absence of any one of the used factors. The decision trees formed are combined and predicted value is found.This value is used to predict the result. The main terminologies in this algorithm are pseudo residuals, shrinkage, decision trees, and prediction value.

## VI. ALGORITHM

Input: training set $\{(x_i, y_i)\}_{i=1}^{n}$ adifferentiableloss function L(y,F(x)), number of iterations *M*.
Algorithm:

I. Initializemodelwithaconstantvalue:
$F_0(x)=argmin\sum^{n}L(y_i,\gamma)$.

II. For$m$=1to$M$:

1. Computeso-called*pseudo-residuals*:
$$r_{im}=-\left[\frac{\partial(L(y_i,F(x_i))}{\partial(F(x)}\right]_{F(x)=F_{m-1}(x)}$$
For i=1,..., n.
2. Fitabaselearner (e.g.tree) $h_m(x)$ topseudo-residuals, i.e. train it using the training set$\{(x_i,y_i)\}^{n}{}_{i=1}$
3. Computemultiplier$\gamma_m$bysolvingthe following one-dimensional optimization problem:
$$\gamma_m=argmin\sum_{i=1}^{n}L(y_i,F_{m-1}(x_i)+\gamma h_m(x_{i)}).$$
4. Updatethe model:
$$F_m(x)=F_{m-1}(x)+\gamma_m h_m(x).$$

III. Output$F_m(x)$.

## VII. CONCLUSION

Inthis research, We have come up withaningenious wayto detect fake accounts on OSNs By using machine learning algorithmsto itsfullextent, we have eliminated the need for manual prediction of a fake account, which needs a lot of human resources and is also a time-consuming process. Existing systems have become obsolete due to the advancement in the creation of fake accounts. The factors that the existing system relayed upon is unstable. In this research, we used stable factors such as engagement rate, artificial activity to increase the accuracy of the prediction.

## VIII. LITERATURESURVEY

Detecting fake accounts in social media has become a tedious problem for many Online Social Networking sites such as Facebook and Instagram. Generally, fake accounts are found using machine learning. Previously used methods to identify fake accounts have become inefficient. In [1], Multiple algorithms like decision tree, logistic regressionand support vector machine algorithms were used for detection. A major drawback of the decision tree algorithmisthatthetreecontainsdatasetsforafeatureandnotfor multiple features. Thus, the models which came after this minimized the number of features as done in [2] where comparing the age entered with their registered mail id and location of the users were used as features. Improvement in creating fake accounts made these methods inefficient in detecting it. Thus, service providers changed their way to predict fake accounts by changing their algorithms as donein [3] where the METIS clustering algorithm was used. This algorithm gets the data and clusters it into different groups which made it easier to separate fake accounts from real accounts. Whereas in [4] Naïve Bayes algorithm is used.The probability for the used features was calculated and is substituted in the naïve Bayes formula and the computed value is checked with a reference value. If the computed value is less than the reference value, then that account is considered to be fake.

## REFERANCES

1. ”Detection of Fake Twitter accounts with Machine LearningAlgorithms” Ilhan aydin,Mehmet sevi, Mehmet umut salur.
2. ”Detection of fake profile in online social networks using MachineLearning” Naman singh, Tushar sharma, Abha Thakral, TanupriyaChoudhury.
3. ”Detecting FakeaccountsonSocialMedia”SarahKhaled,Neamatel tazi,HodaM.O.Mokhtar.
4. ”Twitterfakeaccountdetection”,Buket Ersahin,OzlemAktas, Deniz kilinc,CeyhunAkyol.
5. ” a newheuristic of the decision tree induction” ning li, li zhao, ai-xiachen, qing-wu meng, guo-fang zhang.
6. ”statisticalmachinelearningusedinintegratedanti-spamsystem” peng-feizhang,yu-jiesu,cong wang.
7. ” a study and application on machine learning of artificialintellligence” ming xue, changjun zhu.
8. ”learning-basedroadcrackdetectionusinggradientboostdecision tree”pengsheng,lichen,jingtian.
9. ” verifying the value and veracity of extreme gradient boosteddecision trees on a variety of datasets” aditya gupta, kunal gusain,bhavya popli.
10. ”fakeaccountidentificationinsocialnetworks”loredanacaruccio, domenicodesiato,giuseppepolese.