JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

IDENTIFYING FAKE SOCIAL MEDIA PROFILES USING MACHINE LEARNING TECHNIQUES

[1] Dr. K.N.S. LAKSHMI, [2] DONKADA TEJASWINI
[1] Professor & HOD of Computer Science and Engineering, [2] MCA 2nd year,
[1] Master of Computer Applications,
[1] Sanketika Vidya Parishad Engineering College, Visakhapatnam, India.

ABSTRACT:

In the present day, online social media holds a dominant position in various aspects. There is a significant increase in the number of users utilizing social media platforms every day. The primary advantage of these platforms is the ease of connecting and communicating with people. However, this convenience has also opened doors to potential attacks such as fake identity and false information. A recent survey has revealed that the number of accounts on social media is considerably greater than the actual number of users, indicating a significant rise in fake accounts in recent years. The rise in fake accounts in recent years has presented a significant challenge for online social media providers, as identifying these accounts has proven to be difficult. The presence of fake accounts has led to a surge of false information and advertisements on social media platforms. Traditional methods for identifying fake accounts have become outdated due to the advancements in fake account creation. To combat this issue, new models have been developed using different approaches, such as automatic posts or comments, spreading false information, or spamming advertisements to identify fake accounts efficiently.
The rise in fake account creation has led to the development of various algorithms with different attributes to combat the issue. Previous algorithms such as Naïve Bayes, Support Vector Machine, and Random Forest have become inefficient in detecting fake accounts. To address this problem, a new method was proposed in this research to identify fake accounts accurately. The approach involved the use of a gradient boosting algorithm with decision trees comprising three critical attributes: spam commenting, artificial activity, and engagement rate. By combining machine learning and data science, the proposed method provided an accurate prediction of fake accounts.

KEYWORDS: Data science, Fake account detection, Machine learning, online social media

I.INTRODUCTION

Social media has become an integral part of modern society, serving various purposes such as staying connected with friends and sharing news [1]. The user base of social media platforms has been rapidly increasing [2]. Among these platforms, Instagram has gained significant popularity among users. Instagram has become one of the most widely used social media platforms, with over 1 billion active users [3]. Since the emergence of Instagram, individuals with a substantial following have been referred to as social media influencers. These influencers have now become a preferred choice for businesses to advertise their products and services. Social media has both positive and negative effects on society [4]. The rise of online fraud and the spread of false information are growing concerns associated with social media usage. False information is often disseminated through fake accounts, which have become a significant source of misinformation on social media platforms [5]. Companies that invest substantial amounts of money on social media influencers need to verify whether their followers are genuine or not [6]. Therefore, there is a pressing need for a tool that can accurately detect fake accounts. In this study, we employ machine learning classification algorithms to identify fake accounts based on factors such as engagement rate and artificial activity [7].

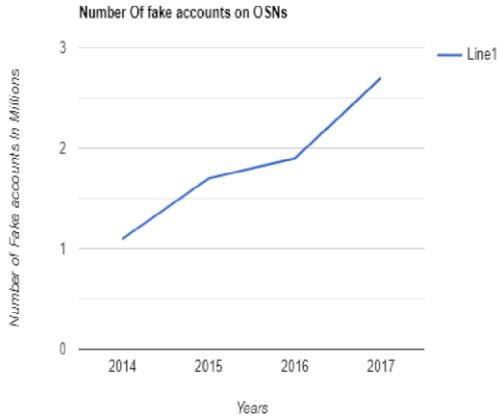


Fig 1.1 Graph Showing increase in number of Fake accounts over the years

II. EXISTING SYSTEM

The current systems utilize a limited number of factors to determine the authenticity of an account, which has a significant impact on the accuracy of the decision-making process [8]. The use of a limited number of factors reduces the precision of the decision-making process considerably [9]. The creation of fake accounts has seen an exceptional improvement, surpassing the capabilities of the current software and applications used for detecting them [10]. As a result, existing methods have become obsolete. The Random Forest algorithm is the most commonly used algorithm in fake account detection applications [11].

The Random Forest algorithm has some limitations, such as its inefficiency in handling categorical variables that have varying numbers of levels [12]. Additionally, increasing the number of trees in the algorithm can negatively impact its time efficiency.

III. PROPOSED SYSTEM

The current system employs the Random Forest algorithm to detect fake accounts, which is effective when all inputs are present and accurate [13]. However, when some inputs are missing, the algorithm's ability to produce accurate output is compromised. To address this challenge, we propose the use of a Gradient Boosting algorithm in the new system [14].

The proposed system for detecting fake accounts employs the Gradient Boosting algorithm, which shares similarities with the Random Forest algorithm in that decision trees are the primary component [15]. However, our approach differs in the methods used to identify fake accounts, which include analysing spam commenting, engagement rate, and artificial activity [16]. These factors are incorporated into decision trees and utilized in the Gradient Boosting algorithm for detecting fake accounts with improved accuracy [17] .

The Gradient Boosting algorithm is advantageous in that it can produce output [18] even when some inputs are missing, which was a significant factor in choosing this algorithm [19] for the proposed system. As a result of utilizing this algorithm, we were able to achieve highly accurate results in detecting fake accounts [20].

IV. DETECTION STRATEGY

Our research defines an account as fake when it fails to meet the minimum engagement rate [21], exhibits artificial activity, or has a history of spam comments.

A. Web Scraper

A Web Scraper is a tool that extracts data from websites [22]. In our system, when a user enters a link to a social media account, we utilize the Outwit Hub Web Scraper to extract the relevant information from the social media site.

Our system extracts various data points from the social media site, including login activity, total likes, total comments, number of posts, number of followings, and number of followers.

B. Calculation of Engagement rate

The engagement rate is a key metric used to measure the level of interaction [23] that a post or story receives on social media. It is calculated as the percentage of the audience that interacts with a post. By comparing the number of interactions to the number of followers, we can evaluate the engagement rate.

Interactions on social media can include likes, comments, and shares. Fake accounts often have a large number of followers but a very low number of interactions. By comparing the engagement rate of popular accounts to semi-popular accounts, it is easier to identify fake accounts. The engagement rate is a crucial metric because a lower rate of audience engagement indicates that the account is likely fake.

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

C. Artificial Activity

When social media activities such as liking, commenting, and sharing occur at an unusually high frequency, they are considered artificial activities, often associated with the use of bots. We also analyse the number of likes, comments, and shares made by an account since its creation to identify fake accounts [24]. If an account has made an excessive amount of likes or comments, it is likely to be a fake account. The term "enormous" refers to a number of likes or comments that is beyond what an average social media user would typically achieve. We also take into account the duration of time the account has been active before making any conclusions. In addition, we consider factors such as incomplete account information and verification status of the mobile number and email.

D. Spam Comments

At this stage, we analyse the comments made from the account in detail. We check if the comments are generic and lack substance, which is a characteristic of bot comments [25]. We also compare the total number of comments made by the user since the creation of the account with the average comments made by users on that particular social media platform. An account may be identified as fake if there is a significant difference between the total number of comments made by the user and the average number of comments made by users on that particular online social network (OSN). Additionally, comments that contain generic or irrelevant content, as well as comments that include links, may also be flagged as spam comments and indicate that the account is fake.

E. Detection of Fake Accounts

The collected data from the website, including login activity, total likes, comments, number of posts, followers, and followings, is combined to compute the engagement rate, artificial activity, and spam comments. Decision trees are created based on these factors using the gradient boosting algorithm. These decision trees are then used to detect fake accounts.

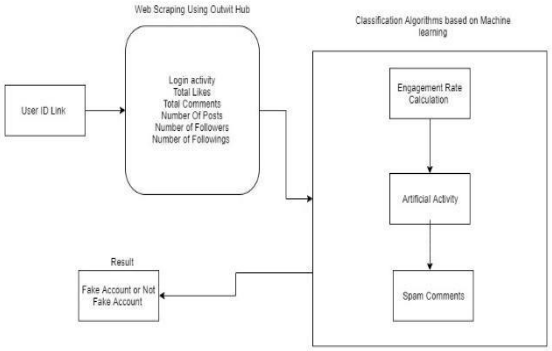


Fig 4.2 Architecture Diagram

V. EVALUATION:

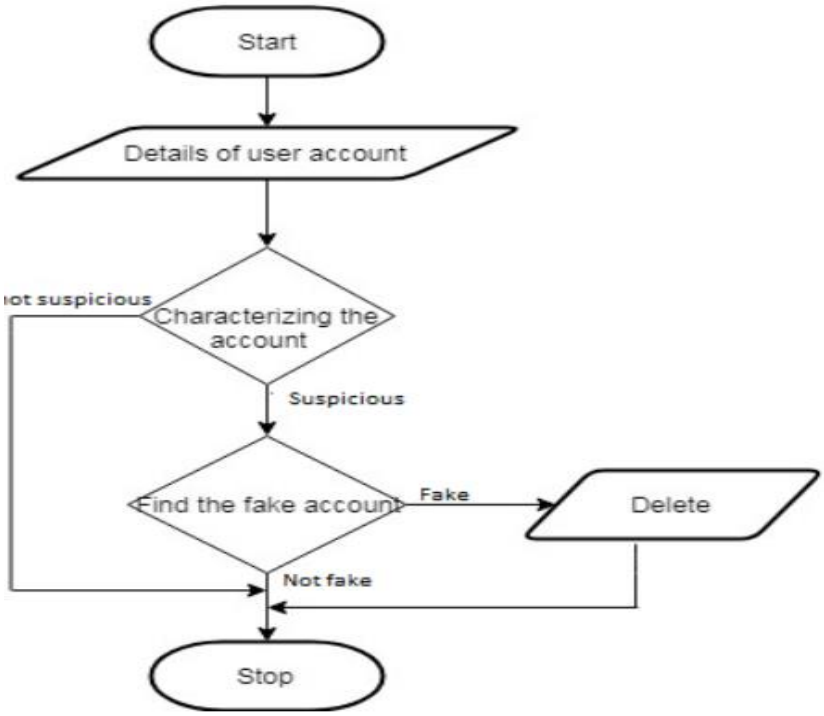
A. Decision Trees

The decision trees are constructed based on the success rate, specifically by examining the values that contain more fake accounts. Initially, the engagement rate is used as the root node for the first tree, with artificial activity and spam comments as child nodes. The third tree is constructed with spam comments as the initial node, followed by artificial activity and engagement rate as subsequent nodes. The second tree is created with artificial activity as the root node, followed by engagement rate and spam comments as subsequent nodes. The first tree is built with engagement rate as the root node and artificial activity and spam comments as child nodes. The third tree is constructed with spam comments as the initial node, followed by artificial activity and engagement rate as subsequent nodes. The second tree is created with artificial activity as the root node, followed by engagement rate and spam comments as subsequent nodes. The first tree is built with engagement rate as the root node and artificial activity and spam comments as child nodes.

B. Gradient boosting

Gradient boosting is a highly effective algorithm for solving classification problems, especially when given accurate input values and a large amount of training data. Its core principle is based on combining multiple weak learners to form a strong rule. One of the key advantages of the gradient boosting algorithm is its ability to make predictions even when one or more input factors are missing. This is because the algorithm combines multiple decision trees to form a strong prediction model. The algorithm relies on several key terms, including pseudo residuals, shrinkage, decision trees, and prediction values. Pseudo residuals refer to the difference between the predicted and actual values, while shrinkage controls the contribution of each tree to the overall prediction model. The decision trees themselves are used to make individual predictions, which are then combined to create the final prediction value.

VI. ALGORITHM



The inputs for the gradient boosting algorithm include a training set, a loss function L(y,F(x)) that is differentiable, and the number of iterations M.

Algorithm:

I. Initialize the model by setting a constant value:

$$F0(x) = \text{argmin}(c)\Sigma L(yi,c)$$

II. For m = 1 to M:

1.Compute the so-called pseudo-residuals:

$$ri = - \partial L(yi,F(xi))/\partial F(xi)$$

2.Fit a base learner (e.g., a decision tree) to the pseudo-residuals, i.e., train it using the training set:

$$hm = \text{argmin}(h)\Sigma L (yi, Fm\text{-}1(xi) + h(xi))$$

3.Compute the multiplier (λ) by solving the following one-dimensional optimization problem:

$$\lambda m = \text{argmin}(\lambda)\Sigma L(yi, Fm\text{-}1(xi) + \lambda hm(xi))$$

4.Update the model:

$$Fm(x) = Fm\text{-}1(x) + \lambda mhm(x)$$

III. Output the final model:

$$F(x) = Fm(x)$$

Note: In this algorithm, L (y,F(x)) represents a differentiable loss function, such as mean squared error or binary cross-entropy, used to measure the difference between the predicted values and the actual values. The training set includes a set of input features (x) and their corresponding output labels (y). The number of iterations (M) is a hyperparameter that determines the maximum number of decision trees that can be added to the model.

A. Random Forest:

Random-forest also known as random-decision-forest is one of the methods that correspond to the category ensemble learning methods. This method is used in machine learning due to its simplicity in solving regression problems as well as classification. Random-forest, unlike the decision tree method, generates multiple decision trees, and the final output is collectively the result of all the decision trees formed.

Similarly, we deployed the random forest [9] method for profile detection. The data is fed to the model and corresponding outputs are obtained. While training, the bootstrap aggregating algorithm is applied for the given set of X = x1, x2.....xn and Y = y1, y2.....yn responses, repeatedly (B times) random sample is selected and fits the trees(fb) to the sample. After training the predictions for a given sample(x') is calculated by the formula specified below:

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x') \qquad (5)$$

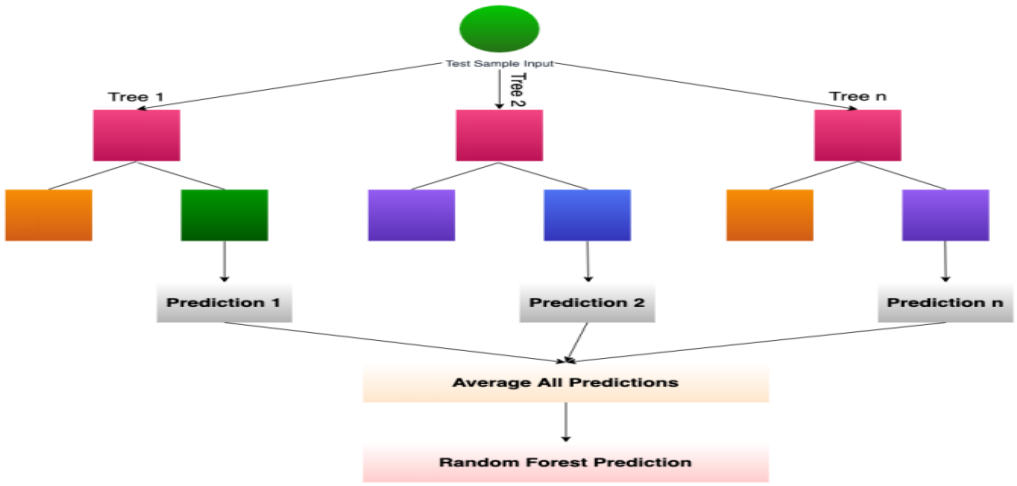


Fig. 4. Random Forest Architecture

XG Boost:

XG Boost is another ensemble learning method used for regression. this implements the stochastic gradient boosting algorithm. Random forest has a drawback, it is efficient only when all inputs are available i.e., there is no missing value. To overcome this, we use a gradient boosting algorithm.

As per the boosting algorithm, firstly, F0(x) is initialized.

$$F_0(x) = argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma) \qquad (6)$$

Then iterative calculation of gradient of loss function takes place

$$r_{im} = -\alpha\left[\frac{\partial L(y_i, F(x_i)}{(F(x_i)}\right] \qquad (7)$$

Finally, the boosted model Fm(x) is defined

$$F_m(x) = F_{m\text{-}1}(x) + \gamma_m h_m(x) \qquad (8)$$

α is the learning rate

γm is the multiplicative factor

Comparing methods:

In the comparison chart below we observe accuracy of different models namely random forest, xg boost, ada boost, and decision tree. The maximum accuracy is achieved by XG boost that equals to 0.996. Further we have decision tree and random forest with approx similar accuracy of 0.99. At last we have ADA boost.

Histogram for accuracy comparison and the ROC curves are as follows:

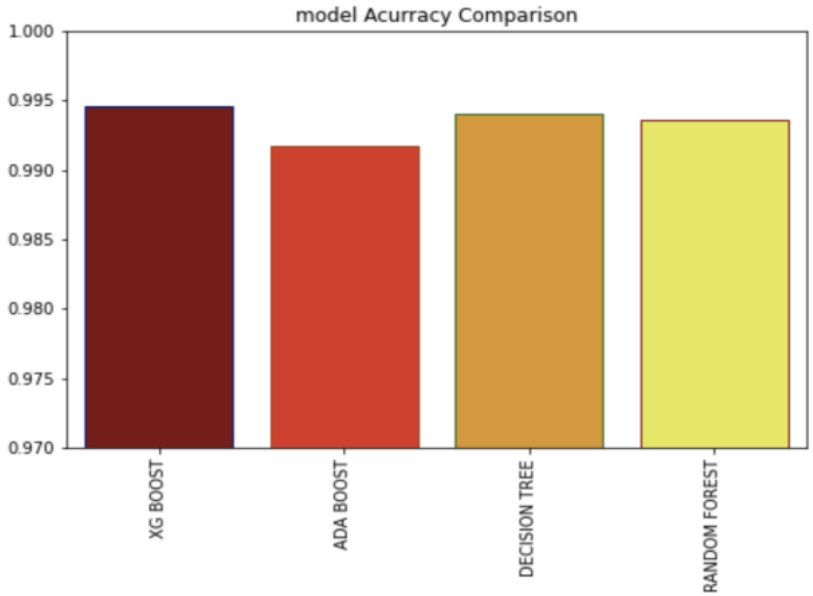


Fig. 8. Accuracy of Different Models

VII. CONCLUSION

This research proposes a novel approach for detecting fake accounts on OSNs through the utilization of machine learning algorithms. By leveraging the full potential of these algorithms, we have been able to eliminate the need for manual identification of fake accounts, which typically requires a significant number of human resources and time. The current systems for detecting fake accounts have become outdated due to the continuous evolution of fake account creation methods. The factors that the existing systems relied on are no longer reliable. To overcome this challenge, we used more stable and reliable factors such as engagement rate and artificial activity in our research. This approach has significantly increased the accuracy of our prediction models.

The identification of fake accounts on social media platforms like Facebook and Instagram has becomes a challenging task. Machine learning is the most commonly used technique to detect such accounts. Subsequent models attempted to overcome this limitation, such as in where features like the user's age, registered email address, and location were used. However, with the advancement in the creation of fake accounts, these methods have become increasingly inefficient in detecting them. As a result, service providers adopted new approaches to detect fake accounts by altering their algorithms, as demonstrated in, where the METIS clustering algorithm was employed. This algorithm clusters the data into various groups, making it easier to distinguish fake accounts from genuine ones. In, the Naïve Bayes algorithm was used instead. The probability of the features utilized was computed and substituted into the Naïve Bayes formula, and the resulting value was compared to a reference value. If the computed value was less than the reference value, the account was deemed fake.

REFERENCES:

[1] A book reference of social media has become an integral part of modern society, serving various purposes such as staying connected with friends and sharing news
https://publications.aap.org/pediatrics/article/127/4/800/65133/The-Impact-of-Social-Media-on-Children-Adolescents?autologincheck=redirected

[2] An article reference of the user base of social media platforms has been rapidly increasing
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0227189

[3] A web reference of Instagram has become one of the most widely used social media platforms, with over 1 billion active users

https://link.springer.com/article/10.17269/s41997-020-00343-0
[4] A book reference of social media has both positive and negative effects on society.
https://www.emerald.com/insight/content/doi/10.1108/09564231311326987/full/html
[5] An article reference of disseminated through fake accounts, which have become a significant source of misinformation on social media platforms
https://www.liebertpub.com/doi/abs/10.1089/big.2020.0062
[6] A web reference of companies that invest substantial amounts of money on social media influencers need to verify whether their followers are genuine or not
https://www.tandfonline.com/doi/abs/10.1080/15252019.2018.1533501
[7] A book reference of machine learning classification algorithms to identify fake accounts based on factors such as engagement rate and artificial activity
https://www.sciencedirect.com/science/article/abs/pii/S0957417421011209
[8] An article reference of the current systems utilizes a limited number of factors to determine the authenticity of an account, which has a significant impact on the accuracy of the decision-making process
https://journals.lww.com/spinejournal/Abstract/2005/10150/A_New_Classification_of_Thoracolumbar_Injuries_.15.aspx
[9] A web reference of the use of a limited number of factors reduces the precision of the decision-making process considerably
https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5915.1999.tb01613.x
[10] A book reference of the creation of fake accounts has seen an exceptional improvement, surpassing the capabilities of the current software and applications used for detecting them
https://www.taylorfrancis.com/books/mono/10.4324/9780080516257
[11] An article reference of The Random Forest algorithm is the most commonly used algorithm in fake account detection applications.
https://www.sciencedirect.com/science/article/abs/pii/S0957417418303579
[12] A web reference of the Random Forest algorithm has some limitations, such as its inefficiency in handling categorical variables that have varying numbers of levels
https://www.sciencedirect.com/science/article/pii/S0268401219302968
[13] A book reference of the current system employs the Random Forest algorithm to detect fake accounts, which is effective when all inputs are present and accurate
https://www.mdpi.com/1424-8220/22/16/5986
[14] An article reference of to address this challenge, we propose the use of a Gradient Boosting algorithm in the new system.
https://www.sciencedirect.com/science/article/abs/pii/S0888327020306397
[15] A web reference of which shares similarities with the Random Forest algorithm in that decision trees are the produce output
https://www.sciencedirect.com/science/article/abs/pii/S0034425704000148
primary component.
[16] A book reference of analysing spam commenting, engagement rate, and artificial activity
https://link.springer.com/chapter/10.1007/978-3-030-71381-2_15
[17] An article reference of algorithm for detecting fake accounts with improved accuracy.
https://link.springer.com/article/10.1007/s11227-018-2641-x

[18] A web reference of The Gradient Boosting algorithm is advantageous in that it can
https://www.sciencedirect.com/science/article/abs/pii/S0034425704000148
[19] A book reference of which was a significant factor in choosing this algorithm
https://dl.acm.org/doi/abs/10.5555/6684
[20] An article reference of able to achieve highly accurate results in detecting fake accounts.
https://www.nature.com/articles/nature07634
[21] A web reference of research defines an account as fake when it fails to meet the minimum engagement rate
https://heinonline.org/HOL/LandingPage?handle=hein.journals/ssbul70&div=7&id=&page=
[22] A book reference of Web Scraper is a tool that extracts data from websites

https://books.google.co.in/books?hl=en&lr=&id=TYtSDwAAQBAJ&oi=fnd&pg=PT30&dq=A+Web+Scraper+is+a+tool+that+extracts+data+from+websites+a+book+reference+&ots=y1BZwGolan&sig=uU9ETzQXsqVohRu1d8NappnlNh8&redir_esc=y#v=onepage&q&f=false
[23] An article reference of engagement rate is a key metric used to measure the level of interaction
https://www.tandfonline.com/doi/abs/10.1080/14719037.2015.1100320
[24] A web reference of analyse the number of likes, comments, and shares made by an account since its creation to identify fake accounts
https://www.inderscienceonline.com/doi/abs/10.1504/IJICS.2020.105181
[25] A book reference of comments are generic and lack substance, which is a characteristic of bot comments
https://www.annualreviews.org/doi/abs/10.1146/annurev-orgpsych-120920-052946

BIBILOGRAPHY:

Dr.K.N.S Lakshmi Currently working as Professor with 16 years of experience in Department of Computer Science and Engineering at Sanketika Vidya Parishad Engineering College, affiliated to Andhra University, accredited by NAAC and currently working as Head of The Department, Published Papers in Various National & International Journals. Her Subjects of interests are Python, Machine Learning, IOT, Data Mining & Data Warehouse.

Donkada Tejaswini is studying her 2nd year, Master of Computer Applications in Sanketika Vidya Parishad Engineering College, affiliated to Andhra University, accredited by NAAC. With her interest in Python, Machine Learning. As a part of academic project, she chooses Identifying Fake Social Media Profiles Using Machine Learning Techniques. A full-fledged project along with code has been submitted for Andhra University as a result of a desire to comprehend, in completion of her MCA.