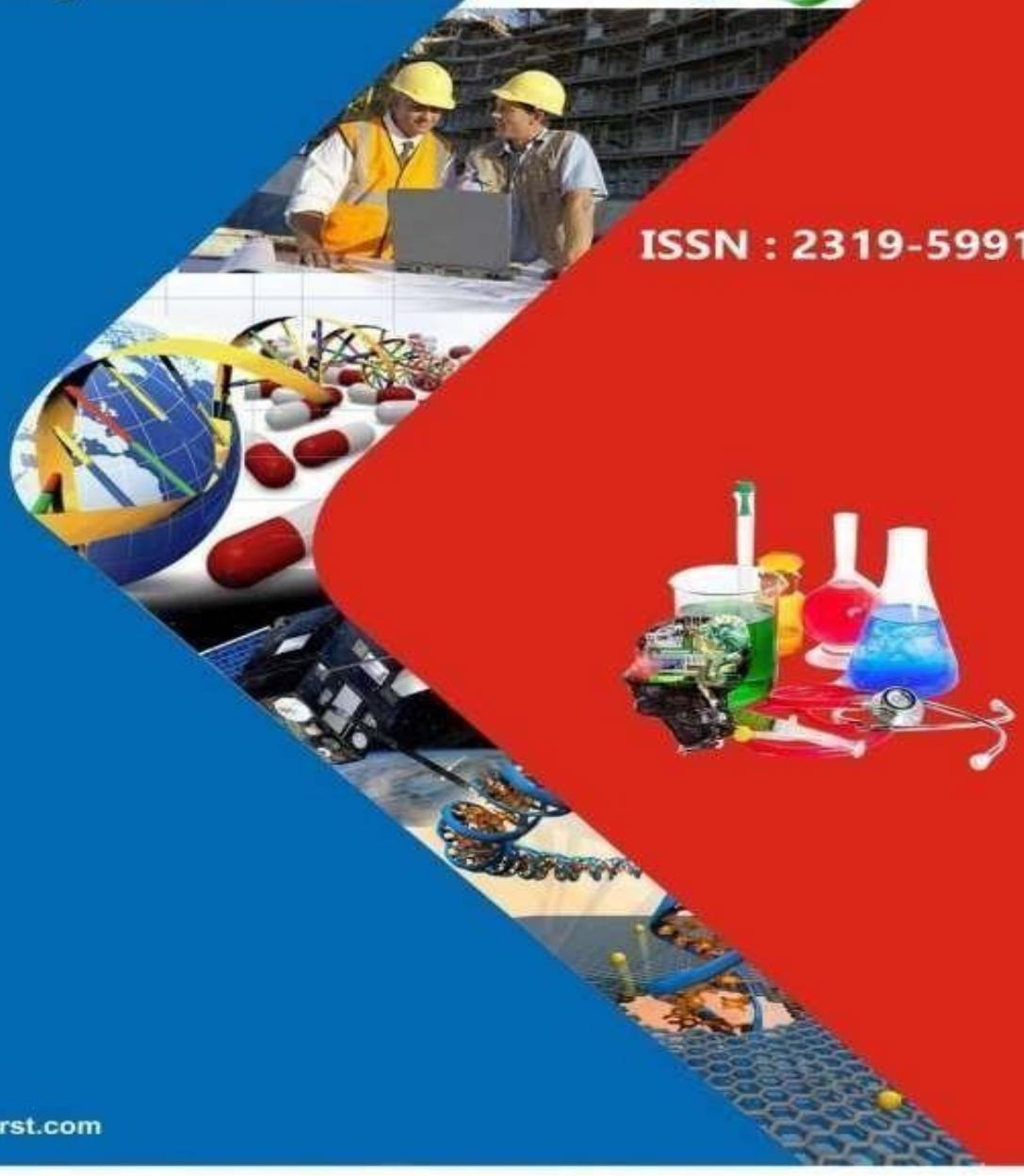
International Journal of Engineering Research and Science & Technology

IJERST

ISSN : 2319-5991



www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

Fake Profiles Identification in Online Social Networks Using Machine Learning and NLP

[1] P.SRINIVASA REDDY, [2] TAMMINA LAKSHMI SIRI MOUNIKA

[1](Assistant Professor), MCA, S.V.K.P & Dr K.S. Raju Arts & Science College

[2]MCA, scholar, S.V.K.P & Dr K.S. Raju Arts & Science College

W.G. District, Andhra Pradesh, psreddy1036@gmail.com

[2]PG, scholar, S.V.K.P & Dr K.S. Raju Arts & Science College(A),

Penugonda, W.G. District, Andhra Pradesh, lakshmisirimounika2001@gmail.com

ABSTRACT

At present social network sites are part of the life for most of the people. Every day several people are creating their profiles on the social network platforms and they are interacting with others independent of the user's location and time. The social network sites not only providing advantages to the users and also provide security issues to the users as well their information. To analyze, who are encouraging threats in social network we need to classify the social networks profiles of the users. From the classification, we can get the genuine profiles and fake profiles on the social networks. Traditionally, we have different classification methods for detecting the fake profiles on the social networks. But, we need to improve the accuracy rate of the fake profile detection in the social networks. In this paper we are proposing Machine learning and Natural language Processing (NLP) techniques to improve the accuracy rate of the fake profiles detection. We can use the Support Vector Machine (SVM) and Naïve Bayes algorithm.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

1. INTRODUCTION

Social networking has end up a well-known recreation within the web at present, attracting hundreds of thousands of users, spending billions of minutes on such services. Online Social network (OSN) services variety from social interactions-based platforms similar to Facebook or MySpace, to understanding dissemination-centric platforms reminiscent of twitter or Google Buzz, to Social interaction characteristic brought to present systems such as Flicker. The opposite hand, enhancing security concerns and protecting the OSN privateness still signify a most important bottleneck and viewed mission. When making use of Social network's (SN's), one of a kind men and women share one-of-a-kind quantities of their private understanding. Having our individual know-how entirely or in part uncovered to the general public, makes us excellent targets for unique types of assaults, the worst of which could be identification theft. Identity theft happens when any individual uses character's expertise for a private attain or purpose. During the earlier years, online identification theft has been a primary problem considering it affected millions of people's worldwide. Victims of identification theft may suffer unique types of penalties; for illustration, they would lose time/cash, get dispatched to reformatory, get their public image ruined, or have their relationships with associates and loved ones damaged. At present, the vast majority of SN's does no longer verifies ordinary users‟ debts and has very susceptible privateness and safety policies. In fact, most SN's applications default their settings to minimal privateness; and consequently, SN's became a best platform for fraud and abuse. Social Networking offerings have facilitated identity theft and Impersonation attacks for serious as good as naive attackers. To make things worse, users are required to furnish correct understanding to set up an account in Social Networking web sites. Easy monitoring of what customers share on-line would lead to catastrophic losses, let alone, if such bills had been hacked. Profile information in online networks will also be static or dynamic. The details which can be supplied with the aid of the person on the time of profile creation is known as static knowledge, the place as the small print that are recounted with the aid of the system within the network is called dynamic knowledge. Static knowledge includes demographic elements of a person and his/her interests and dynamic knowledge includes person runtime habits and locality in the network.

The vast majority of current research depends on static and dynamic data. However this isn't relevant to lots of the social networks, where handiest some of static profiles are seen and dynamic profiles usually are not obvious to the person network. More than a few procedures have been proposed by one of a kind researcher to realize the fake identities and malicious content material in online social networks. Each process had its own deserves and demerits.

The problems involving social networking like privacy, online bullying, misuse, and trolling and many others. Are many of the instances utilized by false profiles on social networking sites. False profiles are the profiles which are not specific i.E. They're the profiles of men and women with false credentials.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

The false Facebook profiles more commonly are indulged in malicious and undesirable activities, causing problems to the social community customers. Individuals create fake profiles for social engineering, online impersonation to defame a man or woman, promoting and campaigning for a character or a crowd of individuals. Facebook has its own security system to guard person credentials from spamming, phishing, and so on. And the equal is often called Facebook Immune system (FIS). The FIS has now not been ready to observe fake profiles created on Facebook via customers to a bigger extent.

2. LITERATURE SURVEY

Fake profile detection in social networks is a critical task due to its implications on user trust, privacy, and security. Leveraging machine learning (ML) and natural language processing (NLP) techniques, researchers have explored various methodologies to identify fraudulent accounts. This literature survey highlights key studies and trends in this domain.

Several studies have proposed ML-based approaches for fake profile detection. Almugren and Hassanein (2018) utilized Random Forest and Support Vector Machine classifiers to analyze profile attributes and user behavior on Twitter. Tare and Bhutada (2019) employed Logistic Regression and Decision Trees on Facebook data, focusing on features extracted from profile attributes and textual content. Safar and Zincir-Heywood (2019) utilized Gradient Boosting and Random Forest on Facebook data, incorporating features related to profile information and network properties.

NLP techniques have also been integral to fake profile identification. Shuet al. (2017) explored sentiment analysis and topic modeling to detect fake news on Twitter, which can be adapted for profile analysis. Venkataraman et al. (2020) proposed a CNN architecture to analyze profile images and textual content for fake profile detection on Twitter, demonstrating the effectiveness of deep learning techniques.

Additionally, studies have examined the integration of ML with network analysis for enhanced detection. Elmi et al. (2020) utilized supervised learning algorithms alongside network analysis techniques to identify fake profiles on LinkedIn, considering both profile attributes and connections.

Despite advancements, challenges persist, including data scarcity and evolving adversarial techniques. Future research directions include exploring novel features, developing more robust algorithms, and addressing ethical considerations.

Several studies have proposed ML-based approaches for fake profile detection. Almugren and Hassanein (2018) utilized Random Forest and Support Vector Machine classifiers to analyze profile attributes and user behavior on Twitter. Tare and Bhutada (2019) employed Logistic Regression and Decision Trees on Facebook data, focusing on features

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

extracted from profile attributes and textual content. Safar and Zincir-Heywood (2019) utilized Gradient Boosting and Random Forest on Facebook data, incorporating features related to profile information and network properties.

NLP techniques have also been integral to fake profile identification. Shuet al. (2017) explored sentiment analysis and topic modeling to detect fake news on Twitter, which can be adapted for profile analysis. Venkataraman et al. (2020) proposed a CNN architecture to analyze profile images and textual content for fake profile detection on Twitter, demonstrating the effectiveness of deep learning techniques.

Additionally, studies have examined the integration of ML with network analysis for enhanced detection. Elmi et al. (2020) utilized supervised learning algorithms alongside network analysis techniques to identify fake profiles on LinkedIn, considering both profile attributes and connections.

Despite advancements, challenges persist, including data scarcity and evolving adversarial techniques. Future research directions include exploring novel features, developing more robust algorithms, and addressing ethical considerations.

checked on full and on selected features set. The proposed method obtained high accuracy.

3. Existing System.

Chai et al awarded on this paper is a proof-of inspiration gain knowledge of. Even though the prototype approach has employed most effective normal systems in normal language processing and human-pc interplay, the results realized from the user trying out are significant. By using comparing this simple prototype approach with a wholly deployed menu procedure, they've discovered that users, principally beginner users, strongly pick the common language dialog-based approach. They have additionally learned that in an ecommerce environment sophistication in dialog administration is most important than the potential to manage complex typical language sentences.

In addition, to provide effortless access to knowledge on ecommerce web sites, natural language dialog-based navigation and menu-pushed
navigation should be intelligently combined to meet person’s one-of-a-kind wants. Not too long ago, they have got accomplished development of a new iteration of the approach that includes enormous enhancements in language processing, dialog administration and information management. They believed that average language informal interfaces present powerful personalized alternatives to conventional menu pushed or search-based interfaces to web sites.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

LinkedIn is greatly preferred through the folks who're in the authentic occupations. With the speedy development of social networks, persons are likely to misuse them for unethical and illegal conducts. Creation of a false profile turns into such adversary outcomes which is intricate to identify without apt research. The current solutions which were virtually developed and theorized to resolve this contention, mainly viewed the traits and the social network ties of the person's social profile .

. The limited publicly available profile data of LinkedIn makes it ineligible in making use of the existing tactics in fake profile identification. For that reason, there is to conduct distinctive study on deciding on systems for fake profile identification in LinkedIn. Shalinda Adikari and Kaushik Dutta researched and identified the minimal set of profile data that are crucial for picking out false profiles in LinkedIn and labeled the appropriate knowledge mining procedure for such project.

Z. Halim et al. Proposed spatio-temporal mining on social network to determine circle of customers concerned in malicious events with the support of latent semantic analysis. Then compare the results comprised of spatio temporal co incidence with that of original organization/ties with in social network, which could be very encouraging as the organization generated by spatio-temporal co-prevalence and actual one are very nearly each other. Once they set the worth of threshold to right level, we develop the number of nodes i.e. Actor so that they are able to get higher photo. Total, scan indicate that Latent Semantic Indexing participate in very good for picking out malicious contents, if the feature set is competently chosen. One obvious quandary of this technique is how users pick their function set and the way rich it's. If the characteristic set is very small then most of the malicious content material will not be traced. However, the bigger person function set, better the performance won.

Disadvantages:

- ❖ The system is not implemented Learning Algorithms like svm, Naive Bayes.
- ❖ The system is not implemented any the problems involving social networking like privacy, online bullying, misuse, and trolling and many others.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

4. PROPOSED SYSTEM

On this paper we presented a machine learning & natural language processing system to observe the false profiles in online social networks. Moreover, we are adding the SVM classifier and naïve bayes algorithm to increase the detection accuracy rate of the fake profiles.

An SVM classifies information by means of finding the exceptional hyper plane that separates all information facets of 1 type from those of the other classification. The best hyper plane for an SVM method that the one with the biggest line between the two classes. An SVM classifies data through discovering the exceptional hyper plane that separates all knowledge facets of one category from those of the other class. The help vectors are the info aspects which are closest to the keeping apart hyper plane.

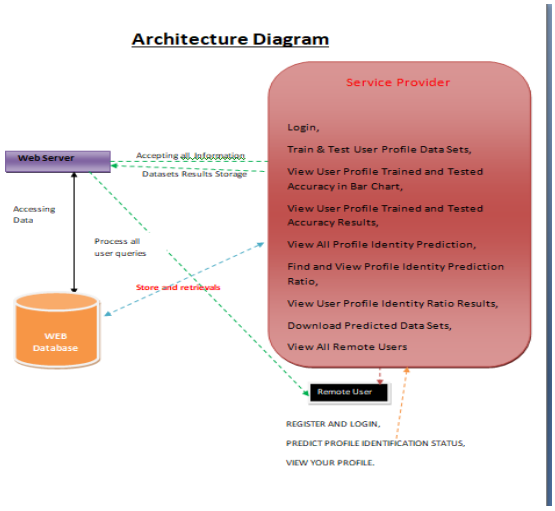
Naive Bayes algorithm is the algorithm that learns the chance of an object with designated features belonging to a unique crew/category. In brief, it's a probabilistic classifier. The Naive Bayes algorithm is called "naive" on account that it makes the belief that the occurrence of a distinct feature is independent of the prevalence of other aspects. For illustration, if we're looking to determine false profiles based on its time, date of publication or posts, language and

Advantages:

In the proposed system, Profile information in online networks will also be static or dynamic. The details which can be supplied with the aid of the person on the time of profile creation is known as static knowledge, the place as the small print that are recounted with the aid of the system within the network is called dynamic knowledge.

In the proposed system, Social Networking offerings have facilitated identity theft and Impersonation attacks for serious as good as naïve attackers.

5. ARCHITECTURE



The above normalization principles were applied to decompose the data in multiple tables there by making the data to be maintained in a consistent state

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

6. Modules:

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as
Login, Train & Test User Profile Data Sets, View User Profile Trained and Tested Accuracy in Bar Chart, View User Profile Trained and Tested Accuracy Results, View All Profile Identity Prediction, Find and View Profile Identity Prediction Ratio, View User Profile Identity Ratio Results, Download Predicted Data Sets, View All Remote Users

View and Authorize Users

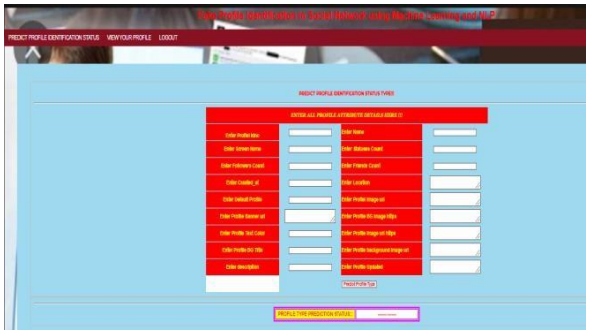
In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

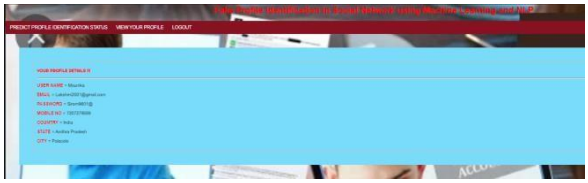
Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN,PREDICT PROFILE IDENTIFICATION STATUS,VIEW YOUR PROFILE.

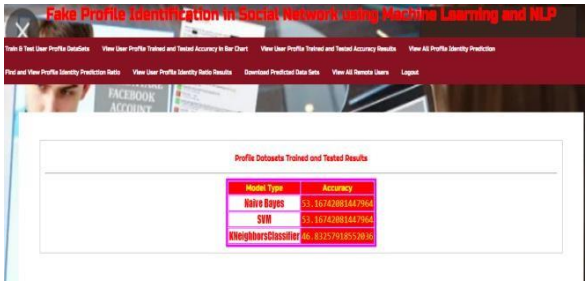
7. SCREENS:











International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024



8. CONCLUSION:

In this paper, we proposed machine learning algorithms along with natural language processing techniques. By using these techniques, we can easily detect the fake profiles from the social network sites. In this paper we took the Face book
Data set to identify the fake profiles. The NLP pre-processing techniques are used to analyze the dataset and machine learning algorithm such as SVM and Naïve Bayes are used to classify the profiles. These learning algorithms are improved the detection accuracy rate in this paper.

9. REFERENCE:

[1] Michael Fire et al. (2012). "Strangers intrusion
detection-detecting spammers and fake profiles in social networks based on topology anomalies." Human Journal 1(1): 26-39.Günther, F. and S. Fritsch (2010). "neuralnet: Training of neural networks." The R Journal 2(1): 30-38

[2] Dr. S. Kannan, Vairaprakash Gurusamy, "Preprocessing Techniques for Text Mining", 05 March 2015.

[3] Shalinda Adikari and Kaushik Dutta, Identifying Fake Profiles in LinkedIn, PACIS 2014 Proceedings, AISeL

[4] Z. Halim, M. Gul, N. ul Hassan, R. Baig, S. Rehman, and F. Naz,"Malicious users' circle detection in social network based on spatiotemporal co-occurrence," in Computer Networks and Information Technology (ICCNIT),2011 International Conference on, July, pp. 35–390.

[5] Liu Y, Gummadi K, Krishnamurthy B, Mislove A," Analyzing Facebook privacy settings: User
expectations vs. reality", in: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference,ACM,pp.61–70.

[6] Mahmood S, Desmedt Y," Poster: preliminary analysis
of google?'s privacy. In: Proceedings of the 18th ACM
conference on computer and communications security",
ACM 2011, pp.809–812.

[7] Stein T, Chen E, Mangla K," Facebook immune
system. In: Proceedings of the 4th workshop on social network systems", ACM 2011, pp

[8] Saeed Abu-Nimeh, T. M. Chen, and O. Alzubi,
"Malicious and Spam Posts in Online Social Networks," Computer, vol.44, no.9, IEEE2011, pp.23–
28.

[9] J. Jiang, C. Wilson, X. Wang, P. Huang, W. Sha, Y.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

Dai, B. Zhao, Understanding latent interactions in
online social networks, in: Proceedings of the 10th
ACM SIGCOMM Conference on Internet Measurement, ACM, 2010, pp. 369–382
[10] Kazienko, P. and K. Musiał (2006). Social capital in online social networks. Knowledge-Based Intelligent
Information and Engineering Systems, Springer