www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

FAKE PROFILE IDENTIFICATION IN SOCIAL NETWORK USING MACHINE LEARNING AND NLP

[1] Dr.A. Avani, [2] T. Bharath Krishna

[1] Associate Professor, Department of Computer Science and Engineering

Anubose Institute of Technology, New Palvoncha-507115,

Bhadradri Kothagudem-Dist-TG

[2] Department of Computer Science and Engineering

Anubose Institute of Technology, New Palvoncha-507115,

Bhadradri Kothagudem-Dist-TG

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.

I.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 Face book 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 naïve 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. The false Face book 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. Face book 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.

II.LITERATURE SURVEY

1.Fake Profile Detection in Social Networks: A Survey Social network platforms have become a major part of people's daily lives, and with that, the problem of fake profiles has grown. Fake profiles are often used to manipulate public opinion, carry out fraud, or perform cyber-attacks. Earlier approaches primarily used rule-based systems and pattern matching to identify fake profiles. However, these methods were limited in accuracy, adaptability, and scalability as they failed to capture evolving behaviour patterns of fake users.

Reference: Chu et al. (2010) pioneered one of the early works on detecting automated accounts (bots) on social networks using machine learning methods. They utilized content-based and behavioural features to classify users as humans or bots. Their approach laid the foundation for further research into automated fake profile detection by demonstrating the utility of machine learning in social media data analysis.

2.**Social Spambots: Detecting Fake Accounts on Twitter Ferrara** et al. (2016) focused on identifying social spambots on Twitter by employing classification algorithms such as Random Forest, SVM, and Decision Trees. Their study analyzed multiple features, such as the number of tweets, frequency of hashtags, followers-to-following ratio, and profile creation dates. The use of feature engineering and social graph analysis proved to be effective, but this approach required significant computational resources.

Limitations: While Ferrara's study improved upon detection, it had difficulties identifying sophisticated fake profiles, which may mimic human behaviour. Additionally, it relied heavily on static profile features that can be easily manipulated by attackers.

3.Combining Textual and Behavioural Features for Fake Profile Detection

A more recent approach combines Natural Language Processing (NLP) with behavioral analysis to detect fake profiles. These methods extract text features such as word usage, sentiment, and grammatical structures from user-generated content (e.g., posts, comments). Chavoshi et al. (2017) proposed the use of time-series analysis along with NLP to detect coordinated behaviour by fake accounts.

Advantages: This study demonstrated that analysing both text content and user behaviour provides a more comprehensive method for identifying fake profiles. Using NLP improved accuracy in identifying profiles that use scripted or copied content.

4.Deep Learning Approaches for Profile Classification

Several researchers have turned to deep learning methods, which are capable of automatically extracting features from raw data. Alomari et al. (2019) introduced a deep learning-based approach to profile detection using recurrent neural networks (RNNs) for analysis user activity patterns. Their method improved accuracy over traditional machine learning models by dynamically learning complex relationships in the data.

Challenges: Despite its success, deep learning approaches are computationally intensive, and model training can be time-consuming. Moreover, they often require large datasets to avoid overfitting.

III.SYSTEM ARCHITECTURE:

The purpose of the design phase is to arrange an answer of the matter such as by the necessity document. This part is that the opening moves in moving the matter domain to the answer domain. The design phase satisfies the requirements of the system. The design of a system is probably the foremost crucial issue warm heartedness the standard of the software package. It's a serious impact on the later part, notably testing and maintenance. The output of this part is that the style of the document. This document is analogous to a blueprint of answer and is employed later throughout implementation, testing and maintenance. The design activity is commonly divided into 2 separate phases **System Design and Detailed Design.**

In system design the main target is on distinguishing the modules, whereas throughout careful style the main target is on planning the logic for every of the modules.

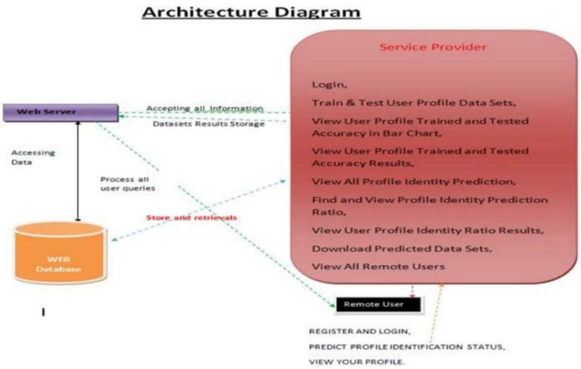
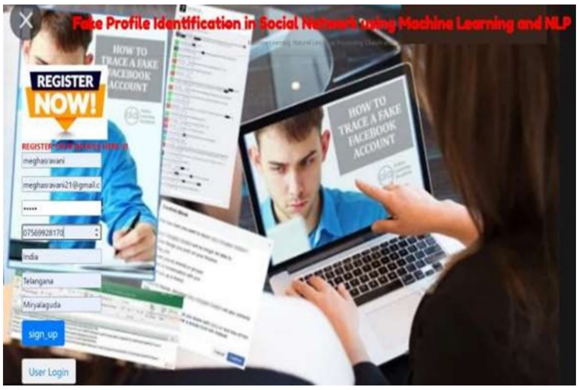

Architecture Diagram
Service Provider
Login,
Train & Test User Profile Data Sets,
View User Profile Trained and Tested Accuracy in Bar Chart,
View User Profile Trained and Tested Accuracy Results,
View All Profile Identity Prediction,
Find and View Profile Identity Prediction Ratio,
View User Profile Identity Ratio Results,
Download Predicted Data Sets,
View All Remote Users
Web Server
Accepting all Information
Datasets Results Storage
Accessing Data
Process all user queries
Store and retrievals
WEB Database
Remote User
REGISTER AND LOGIN,
PREDICT PROFILE IDENTIFICATION STATUS,
VIEW YOUR PROFILE.

figure 3.1 Architecture Diagram

IV .OUTPUT SCREENS


Fake Profile Identification in Social Network using Machine Learning and NLP
REGISTER NOW!
HOW TO TRACE A FAKE FACEBOOK ACCOUNT
sign_up
User Login

meghasravani
meghasravani21@gmail.c
*****
07569928170
India
Telangana
Miryalaguda

Figure No: 4.1 The figure depicts the page where the user have to register using the name, email id ,password ,phone number, country, state, city and then press sign up.

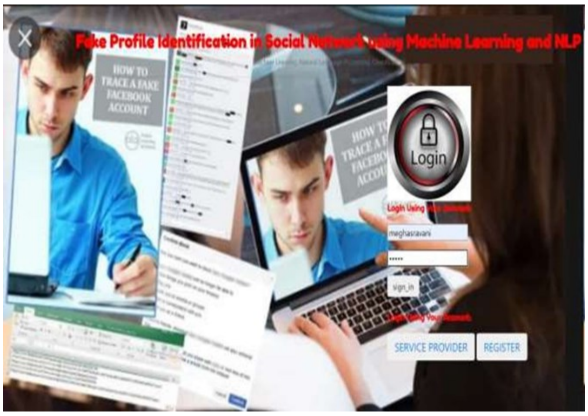
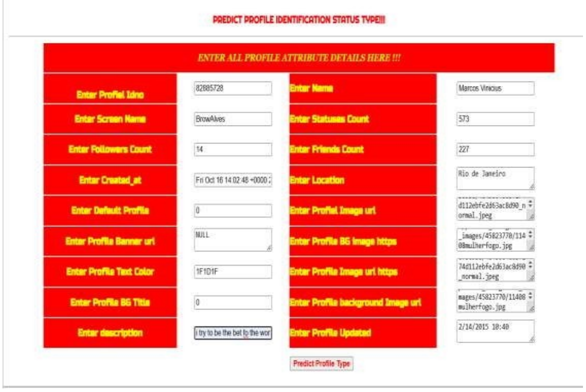

Fake Profile Identification in Social Network using Machine Learning and NLP
HOW TO TRACE A FAKE FACEBOOK ACCOUNT
Login
Login Using Your Account
meghasravani
sign in
SERVICE PROVIDER
REGISTER

Figure No. 4.2: The output figure depicts the login page once the registration is done here user needs to enter the username and password as per registration details.

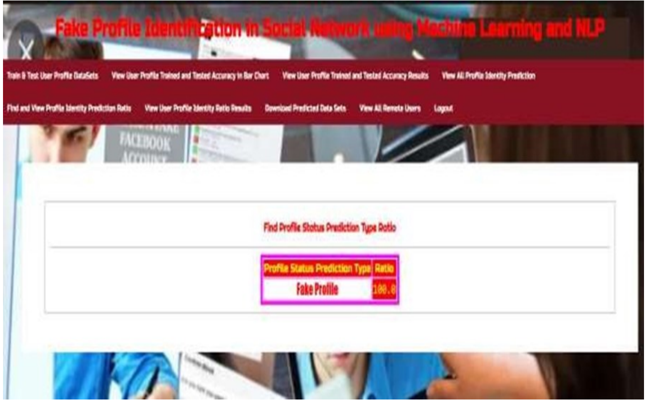


PREDICT PROFILE IDENTIFICATION STATUS TYPE!!!

ENTER ALL PROFILE ATTRIBUTE DETAILS HERE !!!

Enter Profiel Idno	82885728	Enter Name	Marcos Vinicius
Enter Screen Name	BrowAlves	Enter Statuses Count	573
Enter Followers Count	14	Enter Friends Count	227
Enter Created_at	Fri Oct 16 14:02:48 +0000	Enter Location	Rio de Janeiro
Enter Default Profile	0	Enter Profiel Image url	d112ebfe2d63ac8d90_n ormal.jpeg
Enter Profile Banner url	NULL	Enter Profile BG image https	_images/45823770/114 08mulherfogo.jpg
Enter Profile Text Color	1F1D1F	Enter Profile Image url https	74d112ebfe2d63ac8d90 _normal.jpeg
Enter Profile BG Title	0	Enter Profile background Image url	mages/45823770/11408 mulherfogo.jpg
Enter description	i try to be the bet to the wor	Enter Profile Updated	2/14/2015 18:40

Figure No: 4.3: Here details of the profile to be entered regarding the profile which has to be detected accordingly like profileidno, name, screen name etc…..


Fake Profile Identification in Social Network using Machine Learning and NLP
Train & Test User Profile DataSets
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
Logout

Find Profile Status Prediction Type Ratio

Profile Status Prediction Type	Ratio
Fake Profile	100.0

Figure No:4.4 : This shows the status of the profile after detecting by giving the details of the particular profile here it is being detected as Fake Profile.

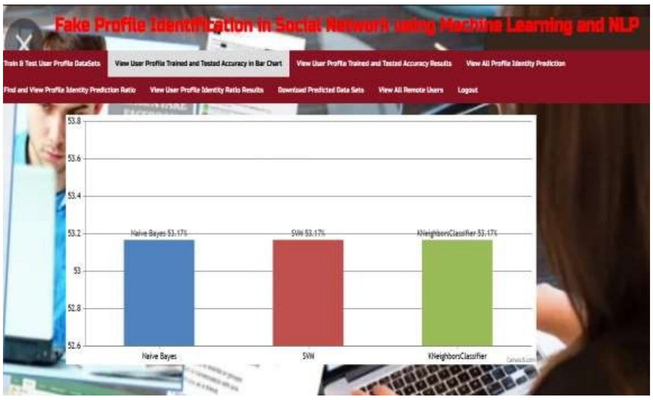


Figure No: 4.5: This shows the bar graph representation of the accuracy rate of each algorithm being used such Naïve Bayes , Support Vector Machine, KNeighborsClassifier.

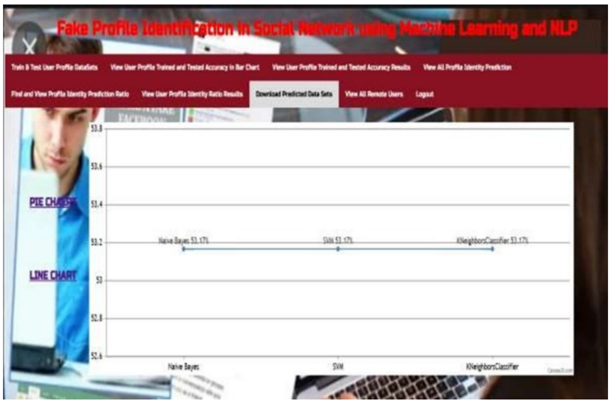
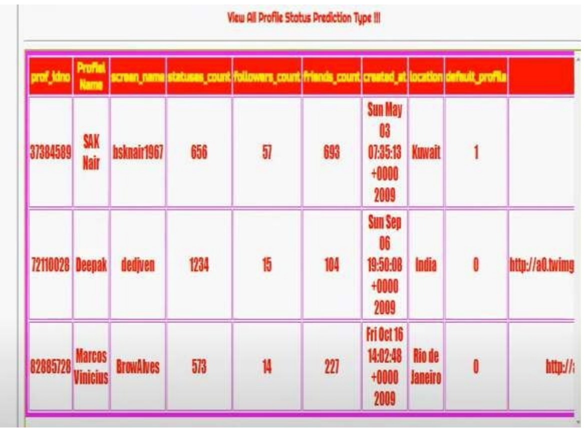


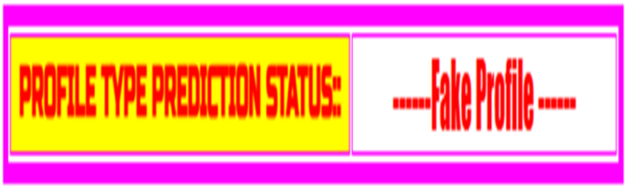
Figure No: 6.6: This shows the line chart of accuracy rate of the algorithms such as Naïve Bayes, Support Vector Machine, KNeighborsClassifier.



View All Profile Status Prediction Type !!!

prof_idno	Profile Name	screen_name	statuses_count	followers_count	friends_count	created_at	location	default_profile	
37384589	SAK Nair	bsknair1967	656	57	693	Sun May 03 07:35:13 +0000 2009	Kuwait	1	
72110028	Deepak	dedjven	1234	15	104	Sun Sep 06 19:50:08 +0000 2009	India	0	http://a0.twimg
82885728	Marcos Vinicius	BrowAlves	573	14	227	Fri Oct 16 14:02:48 +0000 2009	Rio de Janeiro	0	http://

Figure No: 4.7: This output figure depicts the profile prediction status which includes the parameters such as profile id, profile name , status count etc….



PROFILE TYPE PREDICTION STATUS: ----Fake Profile ----

Figure No:4.8: This depicts the status of the profile being detected. So, here the status of the profile is Fake.

V.CONCLUSION

We proposed a novel approach to detect fake profiles on social networking platforms by integrating machine learning (ML) algorithms with natural language processing (NLP) techniques. The fusion of these methodologies enables the effective identification of fraudulent accounts, which is crucial in mitigating misinformation, spamming, and other malicious activities prevalent in online communities.

To demonstrate the practicality of our approach, we utilized a Facebook dataset as the primary source of data. This dataset contains user profiles, which are analyzed to distinguish between genuine and fake accounts. By employing NLP pre-processing techniques, we were able to extract meaningful patterns and clean the data for better analysis. These techniques included tasks like tokenization, stop-word removal, stemming, and lemmatization. This pre-processing step ensures that the data fed into the machine learning models is structured and relevant for classification tasks.

Known for its robustness in handling high-dimensional spaces and its ability to classify data efficiently by finding an optimal hyperplane that separates classes. A probabilistic classifier based on Bayes' theorem, which works well with text-based data and provides quick and accurate predictions. By leveraging these algorithms, the system was trained to recognize distinguishing characteristics of fake profiles, such as suspicious patterns in user behavior, content anomalies, or incomplete profile details. Both classifiers were meticulously trained and tested to ensure high accuracy. The results of our study revealed that integrating these machine learning algorithms with NLP significantly enhanced the detection accuracy rate. The improvement is attributed to the complementary strengths of SVM and Naïve Bayes, along with the thorough data preparation provided by NLP techniques. This methodology not only provides a scalable and efficient solution for fake profile detection but also establishes a foundation for future research in combating online fraud across various social networking platforms.

VI.FUTURE ENHANCEMENTS:

Future Enhancements for the System:

Adapting to Emerging Technologies emerge, the system should be designed in a way that allows easy upgrades and integration of these technologies. This adaptability is crucial for maintaining relevance and competitiveness. For example, cloud technologies, artificial intelligence (AI), and the Internet of Things (IoT) are transforming many industries, and a system that can easily incorporate these advancements will stay ahead of the curve. The system could be modular, with separate components or services that can be updated individually. This allows developers to upgrade the system without overhauling everything. Additionally, building with an API-first approach means new technologies or features can be added through APIs, allowing seamless integration with external services and platforms.

Improving Security Based on Future Challenges Security is always a moving target due to the continuous development of new threats and vulnerabilities. The traditional approach to security may not be sufficient as attacks become more sophisticated, and user expectations regarding security are also higher. Emerging technologies provide opportunities to strengthen security measures Agile Employ Agile development to iterate and continuously improve based on user feedback and changing demands. technologies, and security threats can change over time, it’s important to use an Agile development methodology to accommodate these changes. Agile allows teams to quickly adapt to evolving requirements, incorporate user feedback, and implement incremental improvements.

Scalable Build systems that can scale horizontally and integrate with legacy systems to handle growth and future changes. As systems evolve and user bases grow, scalability becomes a critical factor.

VII.REFERENCES

[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). "neural net: 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 analysisof 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–

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