Indian Journal of Science, Technology & Environment (IJSTE)

A Peer Reviewed, Refereed, Indexed, Quarterly Research journal of Science, Technology & Environment

ISSN : 3108-2017(Online)
3108-1304(Print)
Vol.-1; Issue-2 (Oct.-Dec.) 2025
Page No.- 01-11

https://ijste.gyanvividha.com

Author's :
Dr. Sheena Wahid Khan
DCS & IT, JRN Rajasthan Vidhyapeeth University, Udaipur, India.

Corresponding Author :
Dr. Sheena Wahid Khan
DCS & IT, JRN Rajasthan Vidhyapeeth University, Udaipur, India.

Privacy Preservation In Online Social Network using modeling techniques

Abstract : One essential component of contemporary living is the Internet. This online community brings individuals together by allowing them to share posts, photos, and other types of material. Because communi-cation happens electronically across a variety of media, trust is regularly damaged in all systems that rely on the Internet to function. Although many of these websites include built-in privacy settings, they are limited and do not meet the needs of all users. The proposed study highlights the privacy risks associated with various personally identifiable information published on social media platforms (OSN). This article presents a machine learning method for distinguishing between authentic and fraudulent Instagram accounts with the goal of enhancing social media's authenticity and dependability. The model combines trained and unsupervised anomaly detection tech-niques with access control techniques to protect privacy. The control model updates the properties used to categorize individuals by using Bayesian classifiers to achieve above 95% accuracy, long short-term memory recurrent neural network classifiers to achieve 95.53% accuracy on the receiver operating characteristic curve, and deep neural networks to achieve 95.53% accuracy.

Keywords: Online social network instagram, privacy issues, third party applications, social media privacy.

1. Introduction : Nowadays, user privacy is a big worry with online social networks. However, the protection offered by conventional techniques for machine learning anomaly detection that employ user log data and patterns of behavior is insufficient. Therefore, social network security is use a range of security measures to account for new information and protect user data. More specifically, machine learning techniques may be combined with access control models in the privacy-preservation process. To find unusual people, the models could make use of extra data from the users' profiles. Here, report on a privacy-preserving system integrates models for access control that use anomaly detection methods from supervised and unsupervised machine learning.

The list of characteristics used to categorize individuals is continuously updated by our control model due to the extensive and fine-grained regulations. According to the receiver operating characteristic curve, it has attained 95.53% accuracy with Classifiers for long short-term memory recurrent neural networks and deep neural networks and above 95% accuracy with a Bayesian classifier. According to experimental results, this method performs better than other detection techniques including the Kolmogorov-Smirnov test, isolation forest, principal component analysis, and support vector machines.

The purpose of the Internet's establishment and development is to provide an improved online network for global communication and connectivity. Social network frameworks have been portrayed as a key component of interpersonal communication that offer strong data transmission tools for the sharing of opinions and viewpoints; people rely significantly on these tools as news sources to connect with the network and contribute updates. People can use these tools to investigate various aspects of contemporary situations and share their opinions while receiving feedback; but, if these frameworks become more widely used, political and societal problems will arise.

(Barnes J. A.,1954) first defined informal communities as a social construction consisting of hubs connected by edges that handle at least one specific type of interdependency. To meet the diverse needs of their clients, OSNs come in several varieties. The clients' daily activities hardly ever involve the use of any of the following services: Facebook, Linkedin, Youtube, Instagram, Twitter, and Tumblr. Even if the services provided by different OSNs are meant for varied purposes such as microblogging, systems management, sharing videos, and so on, they all provide a common set of core components. Both the number of OSN clients and the services provided by OSNs are growing at a remarkable rate Through OSNs, users can establish virtual relationships with friends who have been recommended to them as well as with strangers who have similar interests. Users are motivated to disseminate extensive information on social networking platforms for several reasons, including the aspiration to network, cultivate relationships, attain recognition, and conform to social trends.

1.1 Privacy Issues With TPA : The TPA requests the user's attributes in order to offer personalized services to the user. When a user accesses an application for the first time, they must accept the TPA's request to request data sharing. The majority of TPAs operate under the "All or Nothing" model, which forces users to choose between using the service and agreeing to share the attributes indicated as necessary in order to use the application. Through the OSN's APIs, the TPAs are able to access user data. The user has no further control over the data once the TPA has obtained control of it. Sharing exposes the user to a wider range of privacy risks, such as the TPA purposefully or unintentionally disclosing the user's personal information to third-party organizations like advertising agencies, data aggregation brokers, and analytical engines. Therefore, there is a great deal of risk for the user when utilizing the TPA's service.

The user's data is stored by the TPA. Unlike with the user's friend, TPA's relationship with the user is not balanced or transparent (Ali-Eldin 2014). The owner of the application may not always be known to the user. The present "ALL" Because of OR NOTHING's privacy policies, users are forced to grant access to information that may not even be required for the program to function. Social networking sites are accountable for safeguarding user data that is entrusted to them. The customary method involves presenting the user with a acceptance document. It is challenging for users to understand the requested permissions and make an informed sharing choice because the consent pages presented by the TPAs are text-based and have a uniform appearance.

However, once information is shared, there is no way for the OSN site to keep an eye on it. Furthermore, according to (Kelly 2008), the TPAs don't always offer the same degree of privacy protection as Facebook. Any user's personal information may be exposed by malicious programming or poorly written applications. The absence of control mechanisms and ignorance of

the extent of the disclosed attributes are the primary causes of the potential risk.

1.2 Impact of sharing on privacy in OSN : Sharing is the lifeblood of today's online community, where all users, whether deliberate or not, provide some information to the platform. When someone shares intentionally, they share whatever they feel is appropriate to share in a given situation with a group of people. Unintentional sharing involves the collection, aggregation, and correlation of a user's online activities with other publicly accessible data to build a user profile.

In today's web, sharing is highly encouraged. Users who share their attributes run the risk of being exposed to a wide range of privacy threats, even though doing so may be useful. Therefore, services offered by online social media apps should strike a balance between the needs of users' privacy and usefulness. The field of engineering privacy has been the subject of extensive investigation. Still, there is no exact definition for the concept of privacy. According to (Westin 1967) theory of privacy can safeguard themselves by momentarily restricting who else can access their personal information. The right to privacy is the autonomy of personal, organizations, or groups to determine the manner, timing, and extent of personal information shared with other entities.

There is no one essential condition of privacy that one should always aim to achieve, as (Altman 1975) noted in his constant philosophical and dynamic regulation of privacy. Conversely, desired privacy levels change based on the particular circumstance and the nature of the interaction. People constantly modify interpersonal boundaries in order to accomplish the appropriate degree of privacy. According to (Child 2011) Communication Privacy Management (CPM) privacy boundaries range from complete openness to complete secrecy or closeness. An open boundary, which stands for the act of disclosing, indicates a willingness to provide access to personal data by revealing it or granting authority to look at it. Conversely, a closed boundary denotes a process of security and concealment; it denotes information that is private and may not always be available

2. Related work

2.1 Malicious applications : Integrating TPAs with Facebook improves user experience and makes it easier to detect malicious apps. Malicious applications can access user permissions to post content, allowing them to spread malware, spam, and steal personal information (Cluley 2012), which can then be sold to advertisers. TPAs can post content on behalf of users to their friends. Using words like "hottest," "shocking," or "heartbreaking" in a link can lead to clickjacking, malicious links, and phishing attacks (HackTrix, 2012). Malicious applications can be detected using various solutions (Makridakis 2010; Rahman 2016; Egele 2012). (Rahman et al., 2016) developed the FRAppE tool, which accurately detects malicious Facebook applications. Malicious applications are identified based on TPA characteristics and behavior. New tools are being developed to detect malicious applications, but techniques for creating and spreading them are also evolving.

2.2 Spamming Attack : Unwanted messages sent to a large number of consumers via the internet for objectives like phishing, advertising, or malware distribution are referred to as spam. OSN is more susceptible to spamming than email. (Grier 2010; Kaur 2018). On OSN, users are more likely to trust spam messages from friends. Spammers use email lists to send unsolicited emails to a large audience. The TPA's leak of user attributes, including email addresses, simplifies the task at hand. Sharing basic profile attributes is necessary, even if email is not shared. Correlating these attributes with external sources is a simple way to determine email address. Spamming is often used to promote a product and can be combined with other attacks like phishing, posing a significant risk to users. Context-aware spam can be created using information from OSN and external sources to target users and their friends, increasing spam success rates (Brown 2008).

2.3 Linking Attack : OSNs like Facebook require users to register using their real name and email address. OSN services enable users to stay connected with real-life friends in the virtual world. Dating sites like Match.com, Tinder, and others do not require users to reveal their true

identities. In those accounts, users' privacy is protected through the use of pseudo-anonymity. (Acquisti et al., 2015) found that identifying a user on Facebook could lead to a linking attack on other online social networks. Linking profiles in an OSN allows adversaries to gather more details about the user, even if their identity has been confirmed. It might not be appropriate to disclose information that is disclosed in one context in another (Wondracek et al., 2010). Proposed a method to deanonymize users in OSN, revealing a 42% overlap between Facebook and Xing users. User's profile pictures, profile attributes or friends list could be used to identify and link the user in other OSN sites. A study by Liu and Maes found 15% overlap between two major social networking sites (Liu 2005).

3. OSN- Instagram : To address the challenges in identifying bogus accounts, in this proposes a method to increase detection efficiency. Even though you understand how important the relationship is within the similarity statistic on friends' profiles. The overfitting problem needs to be solved using the feature extraction technique. The similarity technique is used to implement the balance between the datasets. The proximity matrix, which is based on graphs, is used to calculate the similarity of profiles. The identification categories of the OSN architecture are discovered by constructing the proposed framework.

3.1 Instagram and its character

i. Robots: Automated accounts, which perform actions like liking posts, following users, or leaving comments, are frequently used to boost engagement metrics or spread spam.

ii. impersonator accounts: Use real user's names, photos, and personal information to create a realistic impression. They are typically designed to mislead followers, steal identities, or harm reputations.

iii. Inactive or incomplete profiles: that have limited profile information and engage in irregular activity. They are frequently used as substitutes for future scams or to artificially increase follower counts.

iv. Scam Profiles: are accounts created to rip off users by promoting bogus products, schemes to become rich quickly, or phishing for personal information. Scammers frequently use unrealistic offers to entice victims.

v. Catfish accounts: profiles that use fictitious identities, often with stolen photos, to trick others into developing relationships. These accounts are frequently used for manipulating emotions and monetary fraud.

3.2 Models used : Discuss about how this study's usage of a machine learning model to identify phony Instagram accounts can contribute to social media becoming a more genuine and reliable platform for all users.

3.2.1 XGBoost : One effective machine-learning technique that can aid in decision-making and data comprehension is called XGBoost. It is a decision tree implementation that uses gradient boosting. Data scientists and researchers throughout the world have used it to enhance their machine-learning systems. The goals of XGBoost are speed, usability, and performance on large datasets.

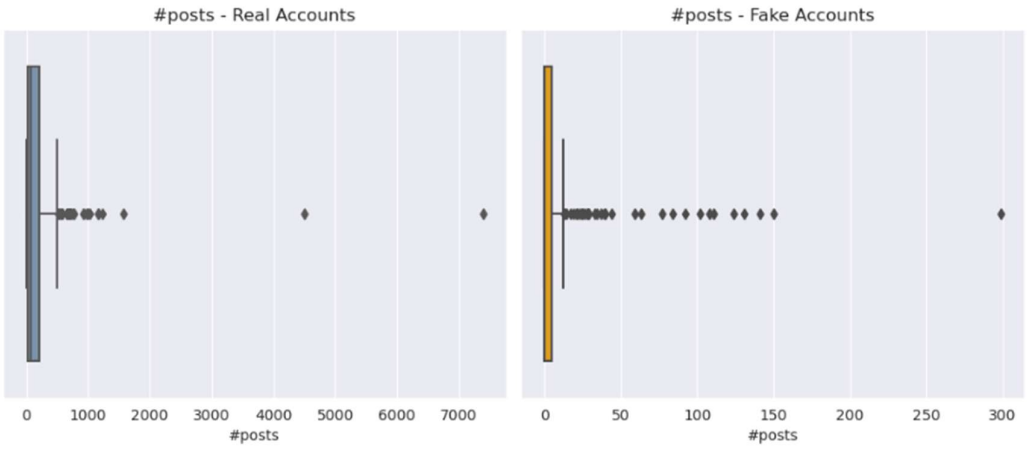
It doesn't need any more configuration after installation because it doesn't need tuning of parameters or optimization.

3.2.2 LightGBM : A distributed, open-source, and highly effective gradient boosting framework is called LightGBM. It is designed to be dependable, scalable, and efficient. It is based on decision trees, which are meant to decrease memory utilization and increase model efficiency. It utilizes several cutting-edge techniques, such as Gradient-based One-Side Sampling (GOSS), which chooses cases with significant gradients during training in order to maximize training time and memory usage. Histogram-based methods are also used by LightGBM to generate trees efficiently. These tactics, along with optimizations like the creation of leaf-wise trees and effective data storage formats, increase LightGBM's effectiveness and give it a leg up on rival gradient boosting frameworks.

3.2.3 CatBoost : CatBoost, or Categorical Boosting is an open-source boosting program developed by Yandex. It is intended use of issues such as regression and classification with a large number of distinct features. CatBoost is a gradient boosting technique that works with both numerical and category data. Category features can be converted to numerical ones without the use of feature encoding techniques like Label Encoder or One-Hot Encoder. It also uses the symmetric weighted quantile sketch (SWQS) algorithm, which reduces overfitting and improves the dataset's overall performance by automatically handling missing values in the dataset.

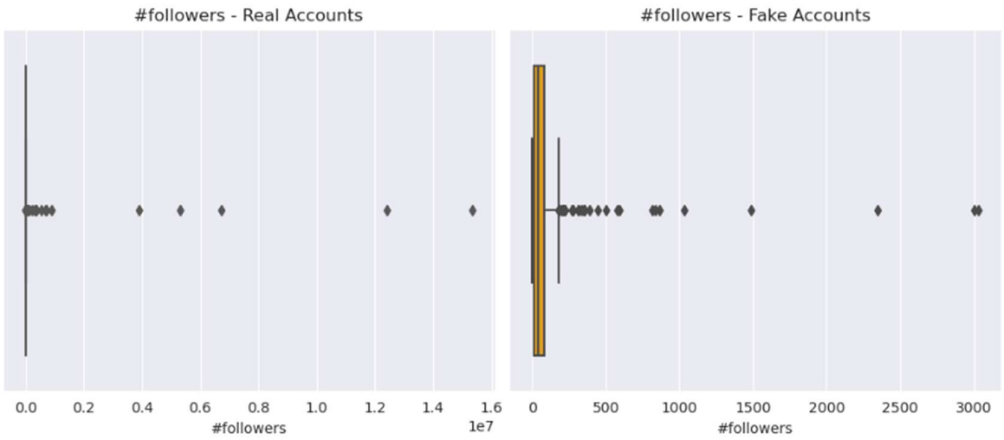
3.2.4 AdaBoost : A machine learning approach called AdaBoost, or Adaptive Boosting, is applied to regression and classification problems. It combines multiple weak or base learners into a strong learner, making it a predictive modeling technique. AdaBoost is an ensemble method in machine learning.

4. Features Used in Models : The features will be used in the sections that follow.



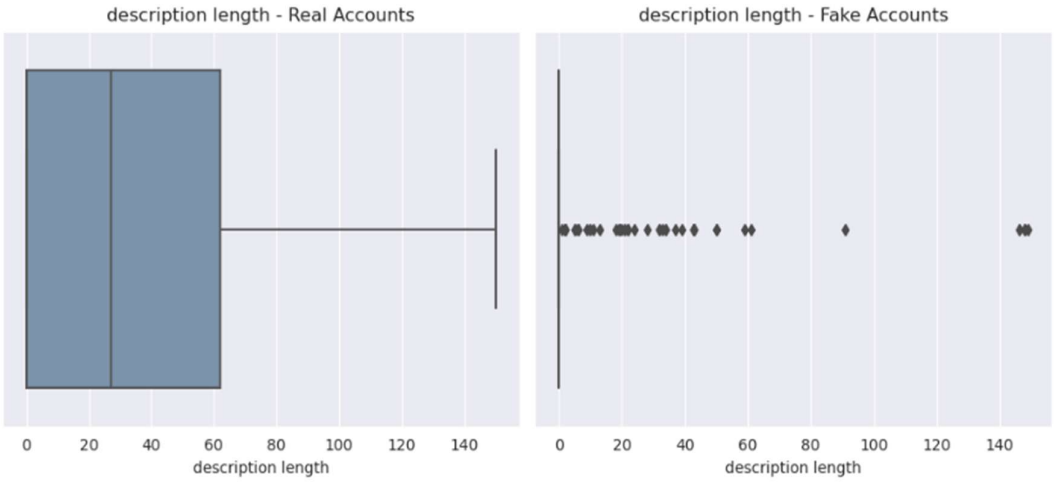
4.1.1 Total number of posts

In general, the majority of postings on both real and fake accounts are around 1,000. It is possible, nonetheless, that real accounts have a significantly greater number of posts than fake accounts when taking into consideration the outliers.

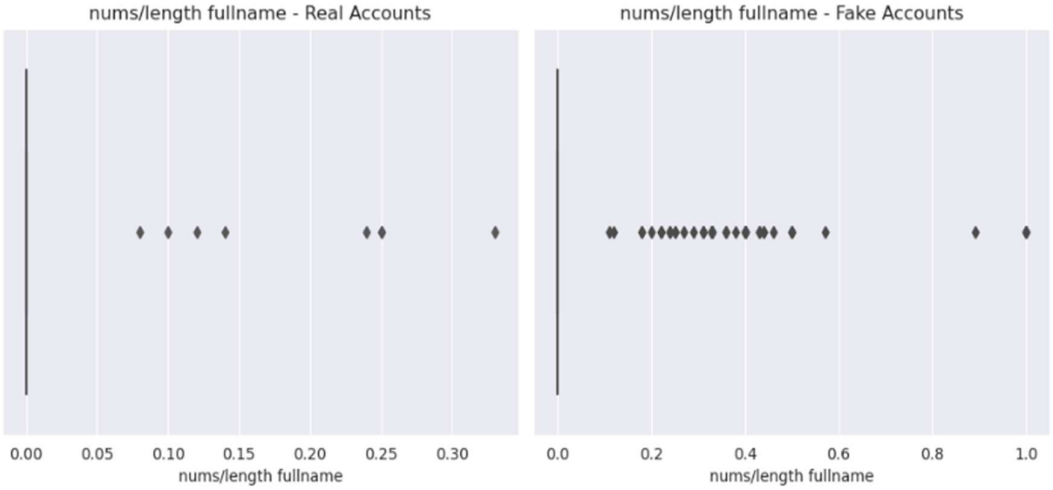


4.1.2 Total number of followers : For real accounts, the x-axis of the number of followers has been shortened using scientific notation. It implies that accounts could have as many as 10,000,000 followers, but fake accounts can only have up to 3,000 followers. Even while It is evident that the outliers of real accounts are greater than those of fake ones, this does not imply that real accounts have more followers on average than false ones. This suggests that the target

variable may have some predictive potential



4.1.3 Bio length in characters : Another fascinating pattern. With very few exceptions, the majority of bios for real accounts are between 0 and 60 characters long. The bulk of fake accounts on the other hand, have blank bios; those with multiple characters are anomalies. This attribute might also have a high degree of predictive potential.



4.1.4 The length of the full name divided by the number of characters

This feature displays, in the above graphic, the proportion of numeric characters in the whole name. Outliers with a high proportion of number characters, however, are more common in fake accounts than in real ones. Whether real or fake, most accounts contain few numbers in their complete names. You can even tell whether an account is fake if it is created entirely of numbers and no letters.

4.2 AUC-ROC score : One significant issue in machine learning is performance measurement. So, the AUC-ROC Curve may be used to solve classification difficulties. Use the AUC (Area Under The Curve) or ROC (Receiver Operating Characteristic) curve to verify or demonstrate the performance of the multi-class classification problem. It is among the most important evaluation factors used to gauge the effectiveness of any classification model It is also known as AUROC (Area Under the Receiver Operating Characteristics). The AUC-ROC score is commonly used to assess the performance of classifiers. Essentially, AUC-ROC scores that are as near to 1.0 as possible are desirable since they show how well the model can distinguish between real and fake accounts.

The True Positive Rate (TPR) is plotted on the y-axis and the False Positive Rate (FPR) on the x-axis in the AUC-ROC Curve. In essence, the FPR indicates how many fake accounts were mistakenly identified, while the TPR indicates what percentage of the positive class that is, fake accounts were correctly categorized. These can be found using the following equations:

TPR = TP ÷ (TP + FN)

FPR = FP ÷ (TN + FP)

AUC-=1

Figure - 4.2.1 AUC-ROC Score = 1

AUC forms above the curve and ranges from 0.5 to 1.0.

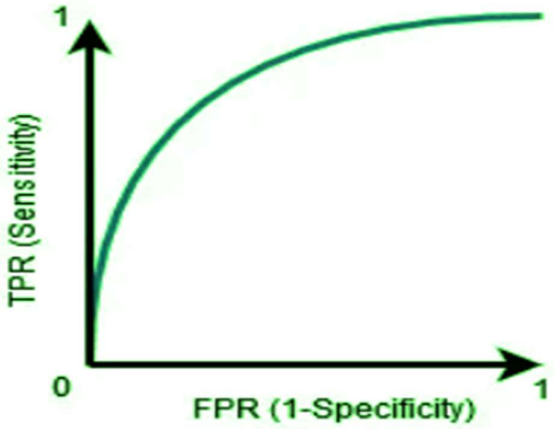


Figure - 4.2.2 0.5 < AUC-ROC Score < 1.0

Furthermore, the AUC is equal to 0.5 in the worst-case situation, when there is a straight line rather than a curve, suggesting that the model is worthless because the classifier performs no better than random guessing.

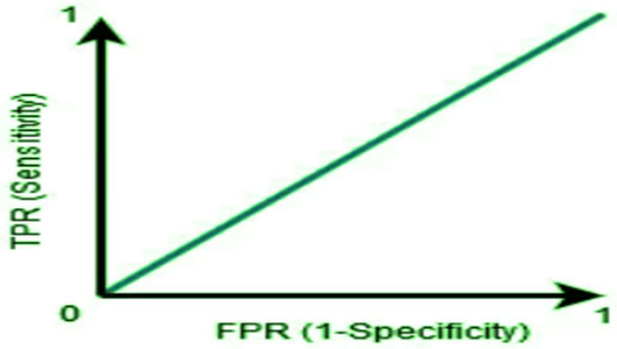


Figure - 4.2.3 AUC-ROC Score = 0.5

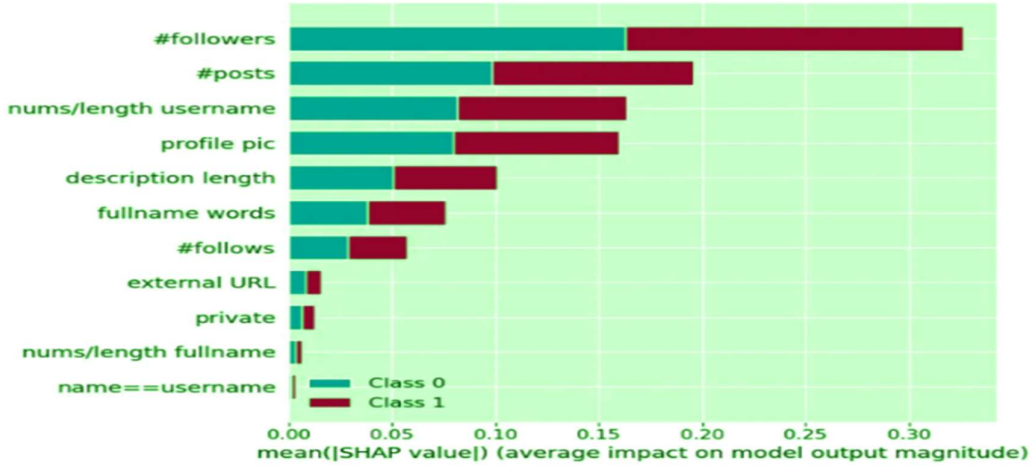


Figure - 4.2.4 Average feature impact on the model's output : As indicated in the exploratory investigation, the most crucial element in identifying whether or not an account was fake, proved its number of followers. The overall count of posts and the proportion of characters in the username that are numbers, and whether or not the account had a profile picture were then found to be significant predictors of the objective variable. In general, every feature seems to have influenced the model in some manner, suggesting that no feature was disregarded as an insufficient indicator for differentiating between real and fake accounts.

5. Feature Engineering : The procedure for developing fresh features based on on preexisting features in the dataset is known as feature engineering. This gives our models more characteristics and data to work with, improving their overall performance and resilience.

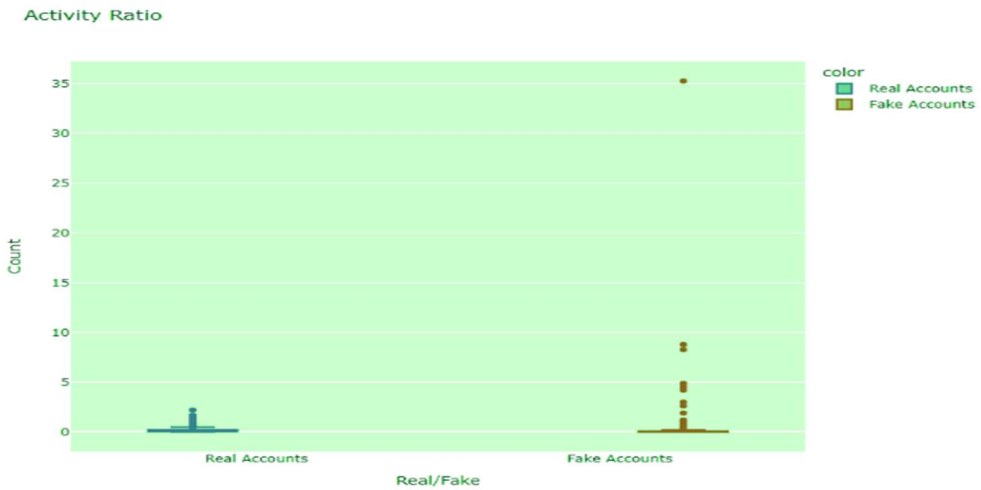
5.1 Activity ratio : This basically compared the quantity of followers with the posting activity of that account. The following is the equation:

Activity ratio = Number of posts ÷ number of followers

believed that most fake accounts have a significantly smaller following than number of users they follow. Add a binary function that reads #followers > #follows, then. Assign a value of one to accounts with more followers than follows, and zero to accounts that follow more than they are followed.

1= Followers> Following

0= Following> Follows



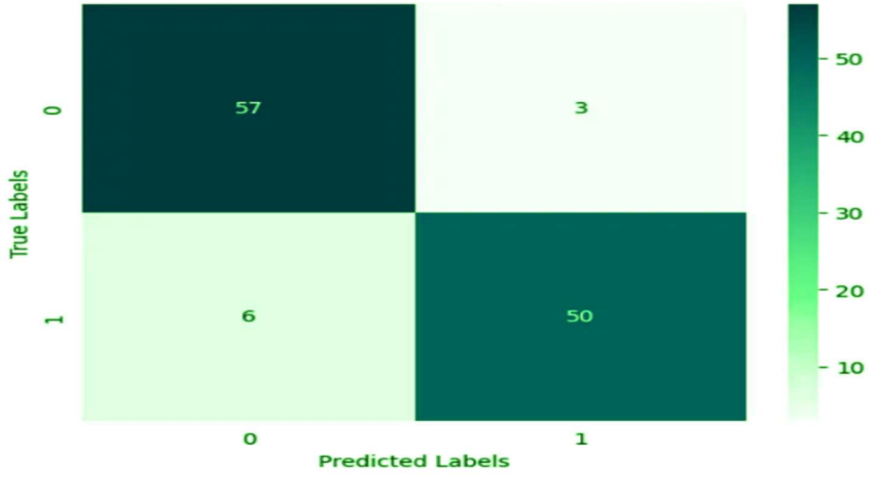
Activity ratio

5.2 Modeling : Now proceed on to the last stage prior to modeling after establishing a baseline and carrying out feature engineering. After that, the program computed the mean and standard deviation. Every feature has a different attribute scale. The standard deviation and mean for every characteristic is computed to determine the degree of diversity in the attribute scales.

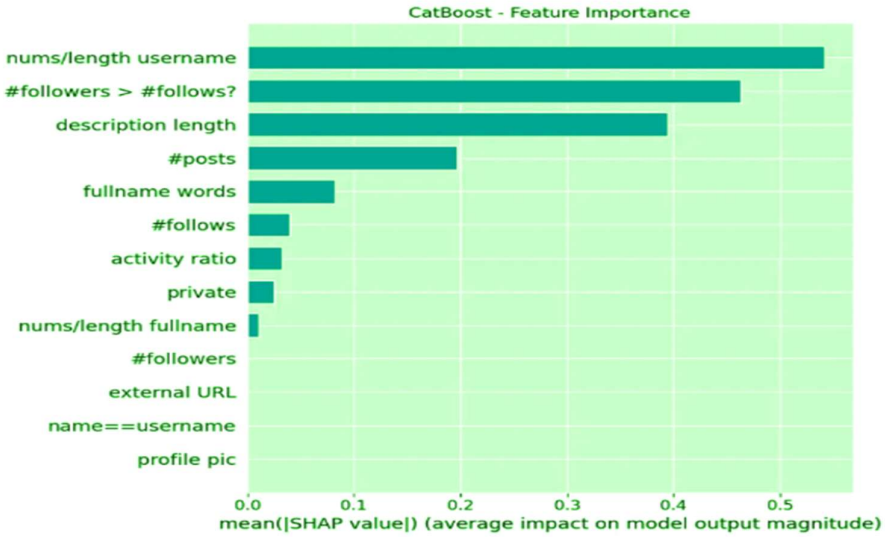
Attributes	Standard Deviation	Mean
profile pic	0.45	0.72
name==username	0.18	0.03
external URL	0.33	0.12
description length	38.13	23.27
fullname words	1.06	1.47
nums/length username	0.21	0.16
nums/length fullname	0.13	0.04
#followers	926256.64	88366.21
#follows	930.82	519.88
#posts	408.69	110.94
private	0.48	0.37
#followers > #follows?	0.50	0.43
activity ratio	1.64	0.28
fake	0.50	0.48

Mean and standard deviation of attributes : The attributes are obviously not on the same scale, which could cause bias in the model. To get around this issue, the models' Pipeline was subjected to the StandardScaler rescaling technique, which made sure that each attribute was scaled consistently.

XGBoost, LightGBM, CatBoost, and AdaBoost are four distinct boosting model are subjected to the modeling.



Confusion matrix of true and predicted labels



Plot of features in the final model : One important component in the final and baseline models was proportion of numeric characters in the username with the CatBoost model giving it the most weight. Additionally, it's very pleasing to note that the feature developed throughout the feature engineering process, together with the activity ratio, came in second for identifying fake accounts, despite not being as important as the other two characteristics mentioned.

6. Conclusion and Future Work : To ensure people's privacy within online social networks (OSNs), this study focuses on extracting patterns and insights from data. The final models output is significantly impacted by this feature alone. The CatBoost model, which scored 0.9416 and accurately identified 90% of phony accounts, was selected to generate final predictions on the test dataset after evaluating a number of boosting techniques and obtaining the greatest AUC-ROC score. After find out the real accounts in the next chapter going to implement a frame work through algorithm which will provide privacy to instagram accounts from TPA. That model will be various forms of access control and to include other attributes for access permission. Additionally, they intend to test generative adversarial network classifiers in adversarial training procedures. Future research must address the malware hierarchical account relationship using an exploration-based methodology. The danger source is determined by the hierarchy, and adaptive loss functions are generated using neural network principles.

References :

1. Barnes, J. A. (1954). Class and committees in a Norwegian island parish. Human relations, 7(1), 39-58.
2. Ali-Eldin, Amr & van den Berg, J. (2014), 'A self-disclosure framework for social mobile applications', New Technologies, Mobility and Security (NTMS), 2014 6th International Conference on, pp 1–5
3. Kelly (2008), 'Identity at risk on Facebook', http://news.bbc.co.uk/2/hi/ programmes/click online/7375772.stm., [Online; accessed 19-June-2015].
4. Altman, I. (1975), 'The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding', ERIC.
5. Westin, A. F. (1967), 'Privacy and freedom atheneum', New York, 7.
6. Child, Jeffrey T & Petronio, S. (2011), 'Unpacking the paradoxes of privacy in cmc relationships: The challenges of blogging and relational communication on the internet', pp 21–40.

7. Cluley, G. (2012), 'The Pink Facebook rogue application and survey scam', https://nakedsecurity.sophos.com/2012/02/27/pink-facebook-surveyscam/, [Accessed 05-feb-2017].
8. HackTrix (2012), 'Stay Away From Malicious Applications', http://www.hacktrix.com/sta y-away-from-malicious-and-rogue-facebookapplications/, [Accessed 05-feb-2017].
9. Makridakis, Andreas & Athanasopoulos, E. . A. S. . A. (2010), 'Understanding the behavior of malicious applications in social networks', IEEE network, 24(5).
10. Rahman, Sazzadur & Huang, T.-K. . M. H. V. . F. M. (2016), 'Detecting malicious facebook applications', IEEE/ACM Trans. Netw., 24(2):773–787.
11. Rahman, Sazzadur & Huang, T.-K. . M. H. V. . F. M. (2016), 'Detecting malicious facebook applications', IEEE/ACM Trans. Netw., 24(2):773–787.
12. Egele, Manuel Moser, A. K. C. . K. E. (2012), 'Pox: Protecting users from malicious facebook applications', Computer Communications, 35(12):1507–1515.
13. Grier, Chris & Thomas, K. . P. V. . Z. M. (2010), '@spam: The underground on 140 characters or less', Proceedings of the 17th ACM Conference on Computer and Communications Security, CCS '10, pp 27–37, New York, NY, USA,ACM.
14. Kaur, Ravneet & Singh, S. . K. H. (2018), 'Rise of spam and compromised accounts in online social networks: A state-of-the-art review of different combating approaches', Journal of Network and Computer Applications.
15. Acquisti, A. and Brandimarte, Laura & Loewenstein, G. (2015), 'Privacy and human behavior in the age of information', Science, volume 347, pp 509–514,American Association for the Advancement of Science.
16. Wondracek, Gilbert & Holz, T. . K. E. . K. C. (2010), 'A practical attack to de-anonymize social network users', Proceedings of the 2010 IEEE Symposium on Security and Privacy, SP '10, pp 223–238, Washington, DC, USA,IEEE Computer Society.
17. Liu, Hugo & Maes, P. (2005), 'Interestmap: Harvesting social network profiles for recommendations', Beyond Personalization-IUI, 56
18. Privacy preservation in online social networks http://hdl.handle.net/10603/303231 (2019) anna university.
19. A novel framework to preserve privacy in online social networks http://hdl.handle.net/10603 /481762 (2022), anna university.

•