G Swathi, 2023, 11:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Fake Profile Identification Using Machine Learning

Asso. Prof. G Swathi, R Vaishnavi, Shaik Noorus Sabiha, P Rakesh Anand, P Nithish Kumar

Sreyas Institute of Engineering and Technology

Abstract- The abstract highlights the enormous user engagement of social net-working sites like Twitter and Facebook and explores their substantial influence on modern digital life. It also highlights the influence that user interactions have on day-to-day life and the growing problem of spammers creating phony personas to disseminate unwanted information. The emphasis is the need for more effective techniques to identify and block phony social media profiles and material. The abstract recognizes the limitations of current machine learning-based techniques, pointing to low accuracy and difficult security maintenance. The suggested method uses supervised learning, especially Extreme Gradient Boosting (Xg boost), to distinguish between authentic and fraudulent profiles more accurately. Furthermore, a web page with a WSGI server is built to help recognize and flag these bogus profiles. Our proposed method gives an accuracy of 99%. The urgent problem of phony profiles and material on social networking sites is being ad-dressed by this strategy.

Keywords- Fake profile, Detection, Machine Learning, social media, Instagram, Internet.

I. INTRODUCTION

Social media like Instagram and Facebook have a big impact on our digital lives, but they are also plagued by fake profiles spreading harmful content. Current methods to tackle this problem are not very accurate.

Our digital lives are greatly influenced by social networking sites like Instagram and Facebook, but they are also plagued by false personas sharing damaging stuff. The current approaches to solving this issue are not particularly effective.

To better identify fraudulent profiles, we apply supervised learning, especially Extreme Gradient Boosting (Xg boost). To help identify and highlight these fakes, we also made a web page. Dealing with the problem of phony profiles on social media is easier this way. The prevalence of false profiles on social media platforms and online groups has grown significantly more problematic in today's digital age. These false personas can be used for a variety of nefarious purposes, such as online harassment, Dis information dissemination, and identity theft. False profiles can be used for a variety of malicious activities, such as distributing false information and indulging in cyberbullying, as well as identity theft and financial fraud. In addition to harming individual users, these false identities erode user confidence and damage the reputation of businesses and organizations by ruining the internet platforms they inhabit.

Researchers and technology businesses have turned to machine learning as a potent method for Fake profile identification due to the severity of this problem. With its ability to examine enormous datasets and identify complex patterns, machine learning presents a possible answer to the enduring problem of fake profile identification. We can improve user security, safeguard data privacy, and uphold the credibility of online communities by utilizing this technology. The goal of this project is to create and put into use machine learning algorithms

and models for identifying false profiles on various internet sites. We seek to develop a robust system capable of discriminating between real user profiles and their fake equivalents through meticulous data collecting, feature engineering, and model training. By accomplishing this, we hope to contribute to the larger initiatives to promote a safer and more reliable online environment.

These are the benefits of this fake profile detection Users can enter data in categories rather than just binary, we handle data scaling and outlier detection, which will make it easier to discover anomalies, and our method is faster and more reliable than others. Our model has a 96% accuracy rate. Additionally, the current approach has the drawbacks of unclear feature relevance and constrained datasets because of privacy issues. For better performance, these problems must be resolved.

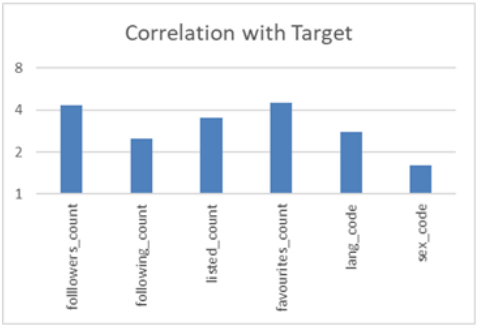


Fig 1 Graph-based Dataset

II. LITERATURE SURVEY

In a paper called "Social Media Malware Detection System", Nambouri et al. [1] have come up with a system to help identify and prevent malicious activities on social media platforms. These days social locales are a genuine risk on the web. To identify fake profiles there are numerous models are proposed. Here, they centered on the Sybil and troll characters utilizing Machine Learning Calculation. Directed Machine Learning Calculation is reused to overcome the issue. Sybil and troll ac-counts utilize a progressed method, the information sets are collected by expansive information blogs at that point put away, in case information becomes pernicious at that point information is cleaned and stored once more, after which cleaned appears the cleaned fake individualities and lost zones are fake individualities. Sometimes recently cleaned store is prepared, and information is put away in a non-relational database for future reference which makes a difference in expelling the fake profile.

Smruthi et al. [2] has utilized a cross-breed show and skin location calculation to identify fake accounts on social sites. The quality of the proposed work is to identify the fake account with tall precision. The result of this proposed work is utilizing 400 blended five Directed Machine Learning Algorithm datasets of fake and genuine accounts. Here there are 200 fake, and 200 genuine accounts that calculate the ex-actness of the Supervised Machine Learning Calculation. On the other hand, the skin location calculation is additionally utilized. Based on the skin exposure rate, the picture is collected for the fake and real accounts calculated. 10-fold cross-validation was applied dataset by employing an administered machine-learning algorithm. Dataset employments 13 highlights of client personality. Here supervised machine learning calculation uses a choice tree classifier having the most noteworthy exactness which is 80% rest of the classifier gave 60% to 80% precision with a 20% mistake rate. The accuracy rate of the KNN calculation is 60% and other classifiers help to extend the precision rate up to 80% such as the Bayes and Choice tree classifier. The skin detection algorithm to begin with was embedded profound learning algorithm for the detection of a picture which picture is human or not. On the off chance that the image may be a human being, at that point apply the skin detection algorithm and calculate the skin uncovered rate. If an image contains more than 13 rate skin ex-posed percentage, it might be considered as a fake profile since most fake profiles is taken a tall rate of skin exposed area or not have any picture. On the off chance that the skin discovery percentage of the picture is tall, accounts are considered as fake accounts.

Tehlan et al. [3] have proposed work to identify spam by exhausting Machine Learning Calculation, Fluffy rationale, and Counterfeit intelligence. This approach overcomes confinement spamming by utilizing the Semi-Directed Learning calculation. Therefore

utilizing fuzzy logic handles the huge information set exceptionally productively and identifies the spam inside a moment, in this paper. it decreases the cost efficiency, time, and utilization of complex programs. The fake accounts on the social location are capable of spamming. The spamming may get to the individual data and start back mail them. Agreeing to a few reports, spam reports spam to contain the account URL.

Srinivas et al. [4] approach Machine Learning and NLP methods to extend the precision rate of fake accounts detection. Through bolster vector and Naïve coves, many problems are expelled on social media, the issue is like privacy issues, cyber bullies, trolls, etc. Most of the time these sorts of malevolent exercises are done by false accounts shown on social destinations. This paper takes the Facebook dataset for the acknowledgment of untrue accounts. Here Machine Learning Calculation besides the NLP preprocessing strategy applies for analyzing the dataset and for classification of the profile utilizing a bolster vector machine classifier that makes a difference in identifying whether the profile is fake or honest to goodness.

Raturi et al. [5] is centered on the engineering of the system. The approach of two concepts of design in the first design will recognize the account's details by using NPL and arrange identifiers. Based on an organized identifier if two or more than two accounts are charged by users they have to be inquired to supply security. In second architecture employments the Bolster Vector Machine BOW (Bag of Words) concept for the recognizable proof of the number of words. The number of words is hurtful, the harmful words collected as the dataset. BOW with predication mode with SVM (Back Vector Machine) connects the dataset and separates the dataset for testing and preparing. It helps to calculate hurtful words from the individual accounts and recognize the wrong account based on the content shown within the accounts at that point send a warning message to give a true demonstration to proceed with the ac-count. On the off chance that the scale is more prominent than three, the client must verify to proceed with the account. The main problem of a social organization isn't giving authentication before distributing the information. In this paper, they used Facebook and Twitter datasets such as Twitter tweets, tags, likes, and posts.

Jan Eloff et al. [6] discuss an engineered feature such as a friend list and the number of follower order of fake profiles for successful detection of a human-created false profile on a social media platform. Numerous characteristics, including name, location, and profile picture, are used here to identify social media accounts. Ac-counts created by humans and bots share similar features, such as names. The ac-counts created by bots are created through an engineered feature. Find the three ma-chine learning modules that are utilized in this paper to successfully identify phony profiles. Entropy is used to identify the engineered features that determine whether an account performed well or not. The entropy that exists there is used to identify the engineered features that determine whether an account performed well or not. Consequently, FI scored 49.7% by accident and on its own. It might forecast 50%, which is not ideal. This is the device The data is trained using a learning model (MLA) with-out any reliance on the set of behavioral data. Occasionally, MLA is obliging in the viral bluff within the extensive dataset.

A technique based on programming structure was suggested by Dhamdhere et al. [7] to determine whether an online presence is authentic or fraudulent. This programmed model has identified profiles using a variety of classifier algorithms, including SVM, Naïve bays, and Decision Trees. Choose the target profile to be tested first, and then the appropriate feature—that is, the feature where the classification method is used. A classifier trained the data using the new dataset using the feature that was extracted. When a target profile is identified by the classifier as being phony, a notification is sent to that profile requesting a genuine identification.

With little effort, the phony profile can be found using this method. Hossam Faris et al. [8] have created a model for the detection of spam profiles. They used 82 feature sets from the Twitter profile dataset to make this detection. Ten binary and basic features are available for usage in the spam profile

classifier. There are two approaches utilized for feature selection: information gain and relief F. Various algorithms for categorization are employed, including Decision Tree, Multilayer Perceptron, k-Nearest Neighbor, and Naïve Bays. The decision tree and naïve Bayes classifier yield a positive result when the feature set is extracted from the presented profile data, which is concerned with the language used by the users' tweets. The outcome shows the words that are suspicious and frequently used, having a significant impact on the accuracy of the detection procedure, and unrelated to the language that users use when tweeting.

Yuval et al. [9] are inventing a method for protecting the privacy of users on online social networks. In this invention, they took the fake and legitimate profile data of the existing social site and then feature is extracted from the pre-determined set of the target fake and legitimate profile, the extract feature set chosen from the friend's list and followers of the target profiles. If examining the relationship of each private node and between the social accounts. Classifiers are applied to other existing features of fake profile, by using trained and non-structured supervised machine learning.

A theoretical methodology for identifying phony accounts is being explained by Suheel Yousuf et al. [10]. The suggested project makes use of machine learning algorithms like ANNs, Decision Trees, Naïve Bayes, and SVM. supervised machine learning algorithm, advanced noise reduction, and data standardization approach data set were used for the detection of fraudulent profiles. The dataset's non-significant at-tributes are found and attribute reduction is performed using the artificial bee colony and ant colony optimization algorithm, which are inspired by nature. The algorithm of the individual support vector machine was implemented for both phony and real profiles. To provide a more accurate prediction in this case, the ensemble classifier is employed.

In their work on the reinforcement concept, Venkatesan et al. [11] demonstrate that the model is capable of successfully identifying bot accounts on its own. The reinforcement model is simply not available on social media, but this model needed a feedback environment that helps to improve the detection process of fake accounts and corrects them effectively. Arif et al. [12] identify spam on social media by using a set of iterative rules and emerging features.

When a new instance of malicious activity is discovered in a filtering technique, the sender is added to a black list. Black listing malicious URL body content on Twitter and detaining bots have been proposed as ways to combat spam [13]. Spammers can use automated and dynamically modified techniques to circumvent the projected methodology, making spam filtering difficult at times. Humans conceal them-selves to avoid being discovered easily. When a profile is blacklisted, all that happens is that someone makes a new one with a fictitious identity [14] as soon as the profile is detected.

In this study, Ebrahimi et al. [15] compared the Naïve Bayes model to the Support Vector Machine model, showing that the one-class model outperforms the binary classification model. Use an outnumbered class to train the one-class Support Vector Machine in this norm.

Gowroju et al.[16-20] experimented on various deep learning techniques to evaluate the performance of prediction using various optimizers. The U Net model using Adam optimizer has performed with a good prediction for predicting the age of the person using Iris biometric.

To determine the objectives, they employ sentiment analysis. The text classifier, which is thought of as a parent algorithm for the text classifier, aids in saving social network space data for improved NLP and SVM implementation. BOW and CNB are both employed by Weka. Weka is a data testing tool that improves the accuracy percentage rate and yields the mode

III. PROPOSED METHODOLOGY

This model was constructed using the XG Boost random forest technique, as well as observable

features from a multi-layered neural network focused on profiles. The extracted features, which were stored in a CSV file, can be easily read by the model. Through training, testing, and analysis, the model will determine if a profile is real or not. This method is effective in detecting false profiles. A training process prepares the model makes the predictions when required and makes them correct if the predictions are wrong. Once the training data can achieve the desired accuracy levels then the training process stops. The accuracy of the model after training may be higher than that of previous studies of a similar nature. Additionally, the design emphasizes a visually appealing framework.

Data preprocessing techniques have been applied to datasets by the proposed model prior to analysis. A method has been used to reduce the number of attributes in datasets that are not significant. The suggested model was trained on a dataset that contained both real and fictitious users using separate supervised machine algorithms. The prediction has been improved through the use of an ensemble classifier.

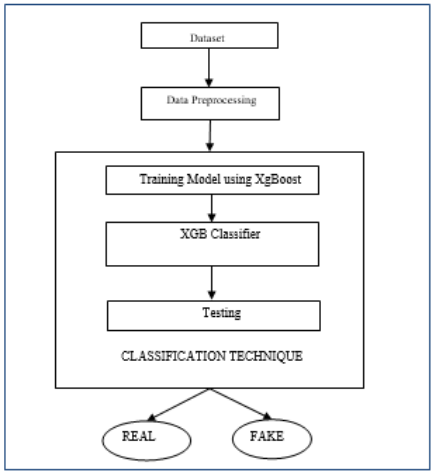


Fig 2 System Architecture

1. Dataset Collection

A dataset is a collection of instances. When working with machine learning methods, we usually need several datasets for different purposes. Training Dataset is a dataset that is fed into our machine-learning algorithm to train the model. Testing Dataset is an example of a dataset that is used to validate the accuracy of our model but not used for training the model. It may also be referred to as the validation dataset. The dataset used in this analysis is the users [11] and Fusers[12] dataset. This dataset consists of 1,482 genuine profiles and 1,338 fraudulent profiles. The data is stored in a CSV format for machine extraction.

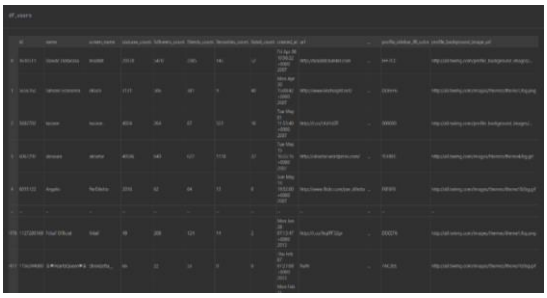


Fig 3 Real User Dataset

2. Data Preprocessing

When using machine learning, it's important to remember that real-world data and images can be incomplete, inconsistent, and have a lot of errors. To make sure the data is ready for the model, there are a few steps you can take.

Data Cleaning

Data cleaning is the process of removing data that has been added or classified incorrectly. Sometimes collected data may have a lot of unwanted data and null values. It is also possible that the data is in an un-structured format. The unwanted data will be replaced by the approximate data. The null values will be filled with some static values. Some collected data sets may have only garbage values. These are some steps in the pre-processing process. Once the data has been cleaned, it is ready for the next step.

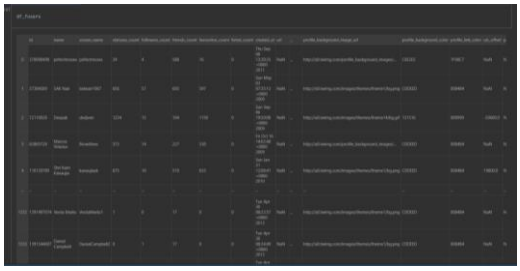


Fig 4 Fake User Dataset

Data Imputations

The majority of Machine Learning frameworks incorporate methods and APIs to facilitate the

balancing or completion of missing data. Generally, techniques include the attribution of missing values using standard deviation, mean, and median values of the data within the specified field.

Over sampling

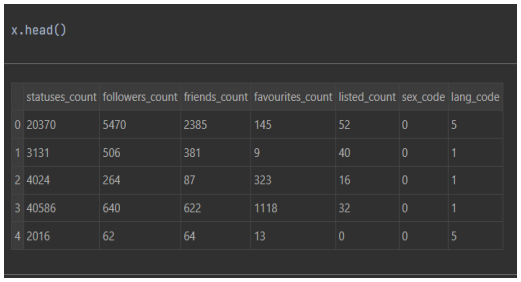
Imbalance in the dataset or bias can be corrected by generating more observations/samples using methods such as repetition, and boot-strap and then connecting them to underrepresented classes.

Data Scaling

Data scaling in fake profile identification using machine learning involves adjusting the range and distribution of your data for better model performance. This helps make features comparable, handle imbalanced data, and improve algorithm performance. It is an important step to ensure accurate and reliable results.

3. Extreme Gradient Boosting (XG Boost) Algorithm

The XG Boost ensemble learning approach is a variation of the random forest approach for regression, which involves the implementation of sub-samples of different stochastic optimization settings. The disadvantage of random forest, however, is that it is most effective when complete inputs are present, or when no missing data is present. To address this issue, the author employs a gradient-boosting strategy.



```
x.head()
```

	statuses_count	followers_count	friends_count	favourites_count	listed_count	sex_code	lang_code
0	20370	5470	2385	145	52	0	5
1	3131	506	381	9	40	0	1
2	4024	264	87	323	16	0	1
3	40586	640	622	1118	32	0	1
4	2016	62	64	13	0	0	5

Fig 5 User's information

The fundamental concept behind gradient boosting lies in constructing a robust rule by combining numerous weak learners. One of its key strengths is its ability to provide accurate predictions even when some of the input factors are missing. It involves aggregating the decision trees and using their collective predictions to make an overall prediction. The crucial terms associated with this algorithm include pseudo residuals, shrinkage, decision trees, and the predicted value.

Training Data

This represents the labeled training dataset, which accounts for 70% of the total dataset.

Validation Data

The validation dataset, which makes up 10% of the dataset.

Test Data

A labeled dataset used for testing, constituting 20% of the total dataset.

Predictions

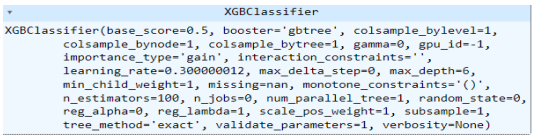
This denotes the predictions generated by the classifiers employed in the process. The validation dataset, "Validation Data," is utilized to validate the predictions made by the classifiers.

Process

Load the training data for each instance within the training data. For every feature matrix supplied to the classifier, including Random Forest, XG Boost (XGB), and Gradient Boosting Machine (GBM). Train the classifier .Evaluate the accuracy and precision of the predictions. Compare the results.

4. XGB Classifier

The XG Boost classifier, when applied to fake profile identification using machine learning, operates as an ensemble learning algorithm. It begins by training on a labeled dataset that includes both genuine and fake profiles. During training, XG Boost constructs an ensemble of decision trees, each serving the purpose of distinguishing between these two classes.



```
XGBClassifier
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=1, gamma=0, gpu_id=-1,
              importance_type='gain', interaction_constraints='',
              learning_rate=0.300000012, max_delta_step=0, max_depth=6,
              min_child_weight=1, missing=nan, monotone_constraints='()',
              n_estimators=100, n_jobs=0, num_parallel_tree=1, random_state=0,
              reg_alpha=0, reg_lambda=1, scale_pos_weight=1, subsample=1,
              tree_method='exact', validate_parameters=1, verbosity=None)
```

Fig 6 XGB Classifier

Importantly, it employs a boosting technique, where subsequent decision trees are trained to focus on

correcting the errors made by their predecessors. This means that instances that were previously misclassified are given higher priority in subsequent trees. Additionally, XG Boost has the capability to assess the importance of different features in the dataset, highlighting those that have the most significant impact on its decision-making process.
XG Boost classifiers are commonly used algorithms that are well-known for their accuracy and efficiency in classification tasks, and they are the foundation of the solution. Authentic and fraudulent profiles are appropriately distinguished in a labeled dataset used to train the model. Using methods such as grid search or random search, hyper parameter tuning further optimizes the model's performance. The model's effectiveness can be fully understood by utilizing evaluation metrics such as ac-curacy, precision, recall, and F1-score.

Understanding the features that set bogus profiles apart requires an interpretation of the model. This is made easier by XG Boost, which shows which features have a significant impact on the decisions made by the model through feature importance scores. Through comprehension of these impactful elements, techniques for identifying fraudulent profiles can be improved. The model can be used in a production set-ting for real-time fake profile identification if it shows promise during the evaluation stage. Periodic updates and ongoing monitoring are necessary to guarantee that the model remains resilient to changing fraudulent activity patterns. In conclusion, using the XG Boost algorithm for fake profile identification not only meets the immediate need for precise classification, but it also offers a framework for continuous enhancements and flexibility in response to new problems.

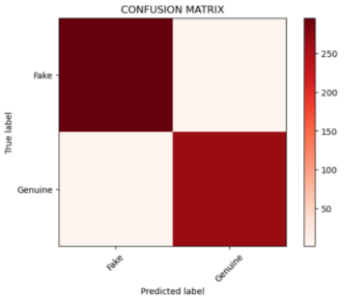


Fig 7 Confusion Matrix

When it comes to making predictions on new, unlabeled data, XG Boost combines the outputs of individual trees to arrive at a final prediction, typically indicating whether a given profile is genuine or fake. Its strength lies in its high accuracy and ability to handle complex data patterns, which proves valuable in tasks such as fake profile identification. Furthermore, the fine-tuning of hyper parameters and cross-validation are often employed to optimize its performance and ensure robust generalization.

IV. RESULTS

One concept that can be used to assess the machine learning model's performance is the confusion matrix.

A confusion matrix can be visualized as a table with four distinct values, such as true positive, false positive, true negative, and false negative, represented in the matrix.

This can be applied to the test dataset where the actual values are known before-hand

- True positives (TP) - the value in which the examples are correctly deter-mined as positive.
- True negatives (TN) - the value in which the examples are correctly deter-mined as negative.
- False positives (FP) - the value in which the examples are negative but are actually determined as positive.
- False negatives (FN) - the value in which the examples are positive but are actually determined as negative.

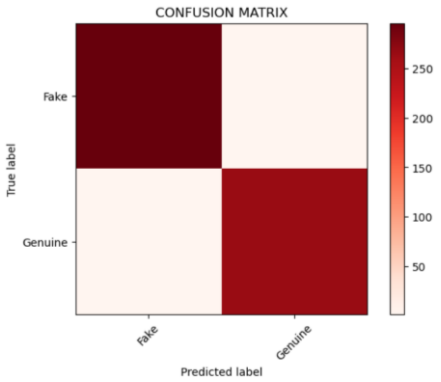


Fig 8 Confusion Matrix

The recall, accuracy, precision, f-measure, and error rate of the classification model are computed by the confusion matrix. The formulas for each of the performance measures listed above are as follows:
TP is equal to TP/(TP+FN).
FP is equal to FP/(FP+TN).
(TP+TN)/(TP+TN+FP+FN) equals accuracy.
TP/(TP+FP) equals precision.
Recall is equal to TP divided by (TP+FN).
Precision * recall/precision + recall equals F-measure.
Error-rate minus accuracy = 1

Evaluation metrics such as accuracy, precision, recall, and F1-score provide a comprehensive assessment of the model's effectiveness in distinguishing between genuine and fake profiles. The interpretability of the XG Boost model, with its feature importance scores, allows for a deeper understanding of the key factors influencing the classification decisions. Once the model demonstrates satisfactory performance, it can be deployed in real-world scenarios to identify fake profiles in online platforms. Regular monitoring and updates are essential to ensure the model's continued effectiveness, as the characteristics of fake profiles may evolve over time.

In summary, the XG Boost algorithm, when applied with a well-prepared dataset and appropriate tuning, can serve as a powerful tool for fake profile identification, offering accuracy, interpretability, and adaptability to changing patterns in fraudulent activity.

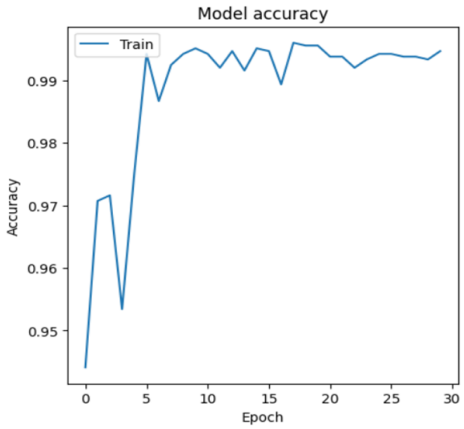


Fig 9 Model Accuracy

The line graph in the image shows the training accuracy of a machine learning model for fake profile identification. The model starts with an accuracy of 0.95, and it gradually increases to 0.99 over 30 epochs. This means that the model is able to correctly identify fake profiles with 99% accuracy after 30 training epochs. This is a very promising result for fake profile identification projects. A model with 99% accuracy would be able to identify a vast majority of fake profiles, which would help to make online platforms safer and more secure. Of course, it is important to note that the model's accuracy may vary in practice, depending on the quality of the training data and the characteristics of the real-world data that it is applied to. However, the results shown in the graph are a good indication that machine learning can be used to effectively identify fake profiles.

Understanding the model's learning curve and possible problems like over fitting or under fitting can be gained from the accuracy graph. Early on, it's possible for training and validation accuracy to increase at the same time. The validation accuracy may even drop if the model over fits, causing the training accuracy to plateau or even worsen. Under fitting, on the other hand, could show up as low accuracies on both sets. The accuracy graph shows how important hyper parameter tuning can be to getting the best possible model performance. The model can identify patterns in the data by varying parameters like regularization factors, tree depths, and learning rates. An optimally tuned model will show a balance between validation and training accuracies, indicating that it can generalize well to new cases.

The accuracy graph can serve as a point of reference for comprehending expected performance in real-world scenarios once the model is deployed. It might be required to conduct routine model updates and monitoring in order to handle changing trends in the characteristics of phony profiles. To sum up, the accuracy graph helps practitioners make well-informed decisions about model deployment and refinement by providing a visual depiction of the model's learning process and generalization abilities.

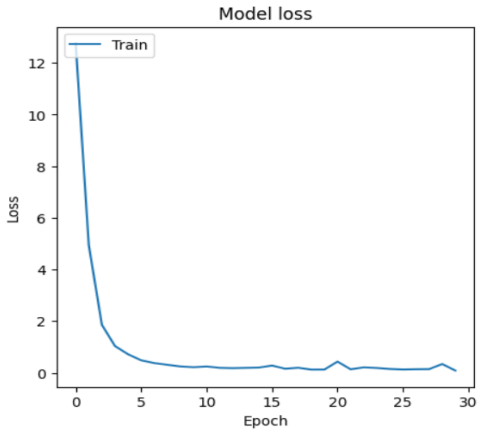


Fig 10 Model Loss

Graph depicting train model loss over time. Loss decreases steadily, indicating model improvement. The loss is a measure of how far off the model's predictions are from the actual labels. In the context of binary classification (distinguishing between real and fake profiles), the loss function is often binary crossentropy. The x-axis typically represents the number of training epochs, where each epoch corresponds to one pass through the entire training dataset .It allows you to observe how the model's performance changes as it iteratively learns from the data. A decreasing training loss indicates that the model is learning, but it's essential to pay attention to validation loss as well to ensure the model generalizes well to new, unseen data. If the training loss is decreasing while the validation loss is increasing, it might indicate over fitting. Adjustments to the model architecture, regularization, or training parameters may be necessary based on the insights gained from the loss graph.

V. CONCLUSION

Finally, we draw the conclusion that further research is needed to locate or identify false user profiles. This work offers an overview and survey of the published and recently submitted research articles. These articles address a variety of security-related topics. The research areas that have been identified include risk assessment, attack-oriented, spamming and fake news distribution, human-created and BOT-based fake accounts, and fake accounts. This paper also reports on the comprehension of social media networks. Additionally, a new data model utilizing data mining and machine learning techniques is proposed by summarizing recent literature. The suggested working model combines methods based on profile attributes and content. By utilizing the benefits of both approaches, the suggested method correctly categorizes the phony accounts. Additionally, the data extraction techniques are completed to implement the proposed data model for identifying the fake user profiles. Furthermore, the various datasets that are available for this task are also reported, and the dataset is finalized for future technique implementation. Additionally, the preprocessing technique is finalized for data processing initially to improve the quality of learning and performance classifiers. Furthermore, a basic synopsis of the suggested data model is provided. The suggested system is put into practice soon, and the results are shared.

Using more complex clustering algorithms, like k-means or hierarchical clustering, is a promising area of research. Although these methods could be beneficial, they pose challenges for large-scale operations: k means might need an excessive number of clusters (i.e., a high value of k) to yield meaningful outcomes, and data clustering might be too labor-intensive for categorizing millions of accounts in online social networks. Applying feature sets from other spam detection models is a crucial path for future work from a modeling standpoint, as it allows for multi-model ensemble prediction. Making the system resilient to adversarial attacks—such as a botnet that expands its features or an attacker who gains experience from mistakes—is another approach.

REFERENCES

1. Nambouri Sravya, Chavana Sai Praneetha, S. Saraswathi," Identify the Human or Bots Twitter Data using Machine Learning Algorithms", International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 03 | Mar 2019 www.irjet.net, e-ISSN: 2395-0056, p-ISSN: 2395-0072.
2. M. Smruthi, N. Harini," A Hybrid Scheme for Detecting Fake Accounts in Facebook", International Journal of Recent Technology and

Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-5S3, February 2019.
3. Tehlan, Pooja, Rosy Madaan, and Komal Kumar Bhatia. "A Spam Detection Mechanism in Social Media using Soft Computing."
4. Rao, P. S., J. Gyani, and G. Narsimha. "Fake profile identification in online social networks using machine learning and NLP." Int. J. Appl. Eng. Res 13.6 (2018): 973-4562.
5. Raturi, Rohit. "Machine learning implementation for identifying fake accounts in the social network." International Journal of Pure and Applied Mathematics 118.20 (2018): 4785-4797. J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," IEEE J. Quantum Electron., submitted for publication.
6. Van Der Walt, Estée, and Jan Eloff. "Using machine learning to detect fake identities: bots vs humans." IEEE Access 6 (2018): 6540-6549.
7. Kulkarni, Sumit Milind, and Vidya Dhamdhere. "Automatic detection of fake profiles in online social networks." Open access international journal of science and engineering 3.1 (2018): 70-73. M. Young, The Techincal Writers Handbook. Mill Valley, CA: University Science, 1989.
8. Ala'M, Al-Zoubi, Ja'far Alqatawna, and Hossam Faris. "Spam profile detection in social net-works based on public features." 2017 8th International Conference on information and Communication Systems (ICICS). IEEE, 2017.
9. Elovici, Yuval, and Gilad Katz. "Method for detecting spammers and fake profiles in social networks." U.S. Patent No. 9,659,185. 23 May 2017.
10. Gurajala, Supraja, et al. "Profile characteristics of fake Twitter accounts." Big Data & Society 3.2 (2016): 2053951716674236.
11. S. Venkatesan, M. Albanese, A. Shah, R. Ganesan, and S. Jajodia, ``Detecting stealthy bot-nets in a resource-constrained environment using reinforcement learning,'' in Proc. Workshop Moving Target Defence, 2017, pp. 75_85.
12. M. H. Arif, J. Li, M. Iqbal, and K. Liu, ``Sentiment analysis and spam detection in short in-formal text using learning classifier systems,'' in Soft Computing. Berlin, Germany: Springer, 2017, pp. 1_11.
13. K. Thomas, C. Grier, D. Song, and V. Paxson, ``Suspended accounts in retrospect: An analy-sis of Twitter spam,'' in Proc. ACM SIGCOMM Conf. Internet Meas. Conf., 2011, pp. 243_258
14. S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, ``Fame for sale: Effi-cient detection of fake Twitter followers,'' Decision Support Syst., vol. 80, pp. 56_71, Dec. 2015
15. M. Ebrahimi, C. Y. Suen, O. Ormandjieva, and A. Krzyzak"`Recognizing predatory chat doc-uments using semisupervised anomaly detection,'' Electron. Imag., vol. 2016, no. 17, pp. 1_9, 2016.
16. Gowroju, Swathi, Sandeep Kumar, and Anshu Ghimire. "Deep Neural Network for Accurate Age Group Prediction through Pupil Using the Optimized UNet Model." Mathematical Prob-lems in Engineering 2022 (2022).
17. Swathi, A., and Sandeep Kumar. "A smart application to detect pupil for small dataset with low illumination." Innovations in Systems and Software Engineering 17 (2021): 29-43.
18. Gowroju, Swathi, and Sandeep Kumar. "Review on secure traditional and machine learning algorithms for age prediction using IRIS image." Multimedia Tools and Applications 81, no. 24 (2022): 35503-35531.
19. Swathi, A. et al. (2023). A Reliable Novel Approach of Bio-Image Processing—Age and Gender Prediction. In: Reddy, K.A., Devi, B.R., George, B., Raju, K.S., Sellathurai, M. (eds) Proceedings of Fourth International Conference on Computer and Communication Technologies. Lecture Notes in Networks and Systems, vol 606. Springer, Singapore. https://doi.org/10.1007/978-981-19-8563-8_31
20. Gowroju, Swathi, and Sandeep Kumar. "Robust deep learning technique: U-net architecture for pupil segmentation." In 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 0609-0613. IEEE, 2020.