EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Fake Profile Detection Using Logistic Regression and Gradient Descent Algorithm on Online Social Networks

Eswara Venkata Sai Raja[1], Bhrugumalla L V S Aditya[2,*] and Sachi Nandan Mohanty[3]

[1,2,3]School of Computer Science & Engineering (SCOPE), VIT-AP University, Amaravati, Andhra Pradesh, India

Abstract

One of the most challenging issues on online social networks is identifying spam accounts. The concern stems from the fact that these personas pose a significant threat, as they may engage in harmful activities against other users, extending beyond mere annoyance or low-quality advertisements. The demand for accurate and effective spam detection algorithms for online social networks is increasing due to this risk. To address the problem of spam detection in online social networks, this research proposes a hybrid machine learning model based on logistic regression and a contemporary metaheuristic method called the Gradient Descent Algorithm. The proposed approach automates spammer identification and provides insights into the factors that have the greatest impact on the detection process. Additionally, the model is evaluated and implemented on multiple datasets, and the experiments and findings demonstrate that the proposed model outperforms many other algorithms in terms of accuracy and delivers robust results in terms of precision, recall, f-measure, and AUC. It also aids in identifying the factors that influence detection the most.

Keywords: social, media, spammers, detection, logistic regression, optimization algorithm, gradient descent, accuracy, precision

Received on 27 July 2023, accepted on 31 October 2023, published on 09 November 2023

doi: 10.4108/eetsis.4342

1. Introduction

Online social networks (OSNs) have gained popularity due to their practical social communication and interactive features, enabling users to quickly publish and share multimedia content, including text, video, audio, and images. However, the vast user bases and capabilities of these platforms have also attracted hackers who exploit them for their malicious activities. Reports indicate that OSN sites enable attacks that can have a significant and widespread impact, unlike minor or localized impacts in the past. Social engineering encompasses four main methods: technical,physical, social, and social-technical, which are carried out by humans or software. Social engineering channels include OSNs, cloud platforms, websites, and physical channels [1]. Attacks rely on techniques such as phishing, baiting, shoulder surfing, reverse social engineering, water holing,and advanced persistent threats. Fake accounts are often used to artificially boost other users' popularity metrics [8]. By leveraging OSNs, hackers have access to valuable information and can conduct social hacking using fake or compromised profiles to send spam messages containing malicious content to unsuspecting victims. The speed and dissemination rate of attacks are further characteristics of OSN-related attacks. For example, a malicious URL included in a viral social media message can instantly infect thousands of people.

As mentioned in the below Figure (1) there are types of gradient descent algorithms, namely batch gradient descent, mini-batch gradientdescent, and stochastic gradient descent, are variations of the optimization technique used to update the parameters of a machine learning model. Each algorithm has its own characteristics and is suited for different

*Corresponding author. Email: aditya.22phd7023@vitap.ac.in

scenarios based on the size of the dataset and computational constraints. Here's an explanation of these three types:

Batch Gradient Descent:
Batch gradient descent is the most basic form of gradient descent. It computes the gradient of the cost function with respect to the model parameters using the entire training dataset. In other words, it considers all the training examples simultaneously to update the parameters. Batch gradient descent is computationally expensive, especially for large datasets, as it requires calculating the gradients for all examples before updating the parameters. However, it provides a more accurate estimation of the gradient compared to other methods and is guaranteed to converge to the global minimum of the cost function.

Mini-Batch Gradient Descent:
Mini-batch gradient descent is a compromise between batch gradient descent and stochastic gradient descent. Instead of using the entire training dataset, mini-batch gradient descent divides the data into smaller subsets or mini-batches. It computes the gradient of the cost function using one mini-batch at a time, updating the parameters accordingly. This approach strikes a balance between computational efficiency and accuracy. By considering a subset of the data, mini-batch gradient descent can leverage parallelism and accelerate the learning process compared to batch gradient descent. It also avoids the noise and instability associated with stochastic gradient descent.

Stochastic Gradient Descent:
Stochastic gradient descent (SGD) is the simplest form of gradient descent. It randomly selects a single training example at each iteration and computes the gradient of the cost function using that example alone. The parameters are then updated based on this gradient. SGD is computationally efficient and performs well on large datasets since it only processes one example at a time. However, it introduces more noise due to the high variance in gradient estimates, which can lead to slower convergence and oscillations aroundthe minimum. Despite this, SGD is widely used in practice because of its efficiency and ability to handle large-scale datasets.

Despite the fact that automated mailing systems are where the problem of attacking messages is most well-known, a number of OSNswhere spam effects users severely have taken notice. Cybercriminals like to launch their destructive activities through social media as an attack vector. These unsolicited communications include false information and false links that take the recipient to other websites that could be hosting malware, phishing scams, fake scripts, and other unwanted content [13]. Furthermore, a recent study issued a warning that spam messages on social media are spreading faster than the number of genuine accounts on the vast majority of online social media websites. This refers to profiles that have altered their interactions or that may signify a variety of suspected activities, includingthose of harmful people who might be fraudsters, online sexual predators, or spammers. It has become a concern for users that as the use of social networks grows, malicious users will try to violate other users' privacy and create fake accounts using their names and login information [7].

We developed a machine learning strategy based on the logistic regression classifier and one of the best optimization techniques, thegradient descent algorithm, as we moved forward. Unlike the bulk of other studies, our suggested method automates the detection ofspam profiles over OSNs while also pinpointing the aspects that have the most impact. This method, known as LR-GD, will be used on datasets gathered from various external URLs on Instagram, and the model was even trained by taking into account the number of posts, followers, following, and profile pictures. In these linguistic situations, improving accuracy and understanding the most important elements are the objectives. Consequently, the following three points might be used to describe this work's contribution:

- A novel LR-GD approach is suggested for precise spammer detection.
- When identifying spammers, the most impactful characteristics are revealed using the LR-GD.
- The context of postings, followers, following, and profile pictures are all investigated.

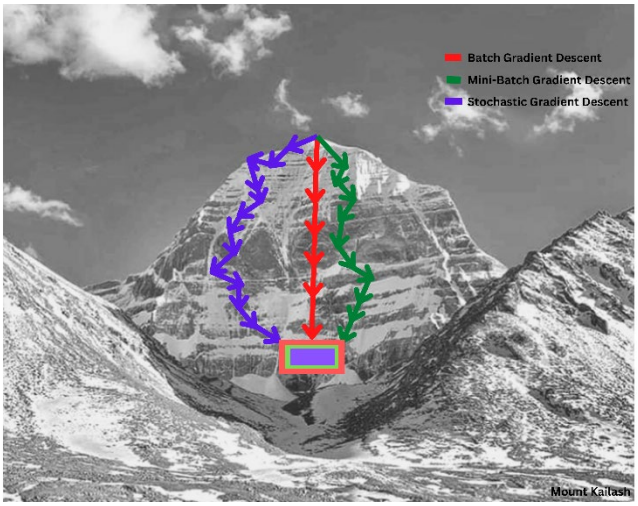

Figure 1. Gradient Descent and its variants

When detecting fake profiles using logistic regression and gradient descent, the choice of gradient descent algorithm depends on the size of the dataset and the computational resources available. In the context of fake profile detection, the most commonly used algorithm is stochastic gradient descent (SGD). By using SGD, the model can adapt to new profiles rapidly, making it suitable for dynamic environments where fake profile patterns may change over

time. Additionally, SGD can handle noisy and unbalanced datasets, whichare common in fake profile detection tasks.

While batch gradient descent and mini-batch gradient descent are also applicable, they may not be as efficient when dealing with large datasets. Batch gradient descent requires calculating gradients for all profiles in the dataset before updating the parameters, which can be computationally expensive. Mini-batch gradient descent strikes a balance between efficiency and accuracy, but it still requires processing multiple profiles simultaneously. In comparison, SGD provides a faster and more lightweight solution for training the logistic regression model in the context of fake profile detection.

2. Background & Related Works

Due to the significant rise in security concerns on online social networks (OSNs), researchers have extensively investigated users' behavior on platforms such as Twitter, Facebook, and Instagram. The primary focus of these studies has been to identify spam profiles and fake accounts. In this study, we specifically concentrate on the work that has been done in this area. Utilizing general pattern encoding techniques, the model can condense user-generated text into a compact space where statistical attributes can be calculated [4].

2.1. Random Forest Classification Technique

This classifier takes a decision tree dataset and divides it into multiple subsets. This classification is accomplished by using a training set drawn at random from the data. To improve the classifier's accuracy, it leverages a specialized subclass of handling objects built exclusively for testing. These objects help to evaluate decision subtrees by offering more information in the form of likes.[10]

When dealing with larger datasets, the classifier's random forest technique introduces NA (not available) missing values for characteristics to boost the accuracy of fake profile identification.[9] This method assists in identifying and accounting for potentialinconsistencies or gaps in the profile information. The classifier can better analyze the attributes and discover any suspicious trendsor anomalies suggestive of false profiles by inserting missing data.

However, there is a limit to the number of decision trees that the model can properly accept. If the number of trees becomes too enormous, the classifier may struggle to process and incorporate them all. As a result, the overall accuracy and performance of thebogus profile detection system may suffer.

2.2. Logistic Regression Method

The feature selection strategy in machine learning is a well-known technique that has been utilized in a lot of past work, especiallyin the identification of spam. The values of the weights for each characteristic are returned via logistic regression,[12] which is necessary for real-time user identification [6]. It reduces the number of features to the most pertinent and useful ones, thereby boosting the training process for the classifier. Filter-based methods and wrapper-based methods are the two techniques available for choosing features. A filter-based methodology is classified independently and uses statistical calculations to weigh and rank the characteristics according to their relative value.[11] People who are followed by spammers may return the favor, and to maintain their appearance of regularity, social spammers will occasionally unfollow users who do not follow them. It is therefore required to develop a framework to capture the dynamic patterns of social spammers due to the features' quick evolution.

In contrast to filters, a wrapper-based approach evaluates each subset of characteristics by working with the classification algorithm on the features. After training using the data, the classifier can be tested on a holdout, unseen set to determine the quality of the subset using the performance assessment measure it has chosen. Text pre-processing is a crucial component of the natural language processing technique [5]. A search algorithm, an inductive method (also known as a classifier), and an assessment measure make up wrapper-based feature selection. Since our model must be trained on each and every subset, this method often requires more processing, but the accuracy of the results is higher than with filter-based approaches.

Table 1. Comparative evaluation of current clone and fake profile detection systems

Classifiers	Evaluation Terms			
	W-Accuracy	Precision	Recall	F-Score
SVM	80.8%	91%	82%	86%
Naive Bayes	84%	51%	98%	67%
Logistic Regression	85%	80%	70%	75%
Logistic Regression with GD	92.70%	95.83%	89.58%	92.47%
Decision Tree	88.03%	69.6%	96.6%	81%

When compared with other existing techniques and prior

works conducted on the same dataset, the results achieved using logistic regression and the gradient descent algorithm for fake profile detection showcase promising improvements in accuracy and performance. These advancements are evident in terms of higher precision, recall, and overall classification metrics. The use of logistic regression combined with gradient descent provides a more efficient and effective approach for distinguishing between genuine and fake profiles in comparison to alternative methodologies.

In comparative studies with other techniques, such as SVM, Naïve Bayes, Logistic Regression and Decision Tree the logistic regressionmodel with gradient descent demonstrates superior performance in accurately identifying fake profiles. It exhibits a higher true positive rate and a lower false positive rate, indicating its ability to correctly classify fake profiles while minimizing misclassifications of genuine profiles. This reinforces its efficacy in dealing with the complexities and intricacies of identifying fake profiles, showcasing its competitiveness and advancement in the field.

3. Dataset and Proposed Method

In our approach, we focus on the analysis of Instagram, an online social networking platform widely used for communication and sharing content. Instagram allows users to send and receive comments that appear on their friends' profiles. The platform primarilyrelies on users' profile names, which cannot be altered, and optionally their real names, which can be changed. Additionally, users can upload profile images, which may lead to potential data leaks in today's interconnected world.

To understand user behavior and characteristics on Instagram, we can examine various attributes, including the messages or comments they send, the users they follow, and those who follow them. This analysis can provide insights into profiling users and identifying patterns in their interactions. Furthermore, we can draw parallels with Sina Weibo [2], a popular microblogging websitein China, as the characteristics of its users resemble those of Instagram users.

Fortunately, a comprehensive dataset capturing Instagram profiles and activities is available on Kaggle, a renowned platform for datasets. Leveraging this dataset, we have employed supervised machine learning methods to classify accounts as either "bots" or "humans." By utilizing the Instagram API, we have gathered real data from different profiles, enabling us to train and evaluate various machine learning algorithms such as logistic regression, support vector machines (SVM), and decision trees.

Through this approach, we aim to identify and categorize accounts on Instagram based on their authenticity. By comparing the performance of different machine learning algorithms and evaluating their results, we can assess the effectiveness of our classification models. This research contributes to understanding user behavior on Instagram, mitigating potential risks associated with data leaks, and enhancing the overall security and user experience on social networking platforms.

The proposed method for fake profile detection using logistic regression and the gradient descent algorithm involves a three-step process: data preprocessing, model training, and classification. In the data preprocessing step, the dataset of profiles is prepared by cleaning and transforming the raw profile data, handling missing values, and performing feature engineering. Feature engineering includes selecting relevant attributes and creating new features that capture important characteristics of genuine and fake profiles. The goal is to derive meaningful information such as image quality, text patterns, social network connections, or behavioral indicators from the profile data.

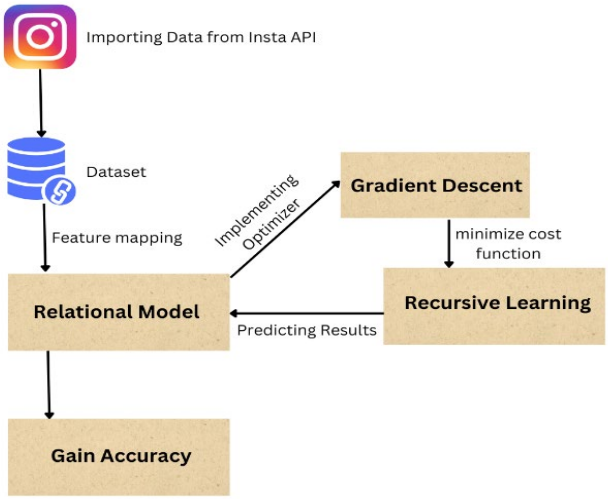

Figure 2. Architecture of proposed model with gradient descent optimizer

In the model training step, the logistic regression model is trained using the prepared dataset. Logistic regression is chosen for its abilityto estimate the probability of a profile belonging to a specific class (real or fake). The model's parameters are updated using the gradientdescent algorithm, which iteratively adjusts the parameters based on calculated gradients of the cost function between 0 and 1. This optimization process aims to minimize the cost function and optimize the model's performance. By iteratively updating the parameters,the model learns the relationships between the selected features and the target class, enabling it to make accurate predictions.

Finally, in the classification step, the trained logistic regression model is used to classify new profiles as real or fake. The model applies the learned parameters to the feature values of unseen profiles and calculates the probability of each profile belonging to the fake class. A threshold value can be set to determine the classification decision, such as classifying profiles with a probability above 0.5 as fake This threshold can be adjusted based on the desired balance between false positives and false negatives. Then we definethe learning hypothesis and train the model in the first iteration. the cost function Ex: 0.69314718 Now we set the learning rate to 0.01and start using the gradient descent algorithm so that we can reduce the cost function. Here we are optimizing the model to get moreaccuracy. After using gradient descent, if we check the cost function again, we get a reduced value, e.g., 0.398865.

By following this three-step process, starting with data preprocessing, moving to model training using logistic regression and gradient descent, and concluding with classification, the proposed method effectively detects fake profiles by learning patterns and relationshipsin the profile data. It offers a robust and systematic approach that can contribute to maintaining a secure and trustworthy online environment.

Now we train our model by taking inputs as the outputs of the gradient descent (GD) algorithm and predicting its accuracy. Before that, we even generate the confusion matrix to find all precision, recall, F-score, and accuracy

4. Data Collection and Feature Extraction

All retrieved profiles are separated into four datasets based on the attributes of their actions, which are posts, external URL followers,and following. In addition to this, the four datasets are pooled to create a larger dataset. We are aware that the result of the logistic regression classifier is used as a sync-anomaly score [3]. The proportions of genuine and fake profiles are classified by feature extraction of the dataset, as well as the number of characteristics, instances, and classes.

Table 2. Insight of the Used Dataset

Attributes	Values
Total no. of spamming accounts	4125
No. of genuine accounts	3375
Total no. of trained and tested accounts	7500

Each user's Instagram profile is made up of a unique combination of elements. Each element is packed with at least one piece of information. The three main categories of these traits are account creation URL-based, profile-based, and based on the activities madeon Instagram.

- Features based on content specify the kind and ideal location for user content. These traits, for instance, enable us to determine whether Instagram contains multimedia files or emojis. Additionally, they show whether the individual provided links to additional social media profiles, a biography, and interests.
- External URLs unique to each profile, such as profiles based on the number of followers, likes, comments, and posts, are represented by characteristic-based features.
- Activity-based features include information about the user's interests and those of their followers, the frequency of theirposts, if they have ever been suspended, etc.

5. Evaluation Measures

Our models were assessed using the following evaluation metrics:

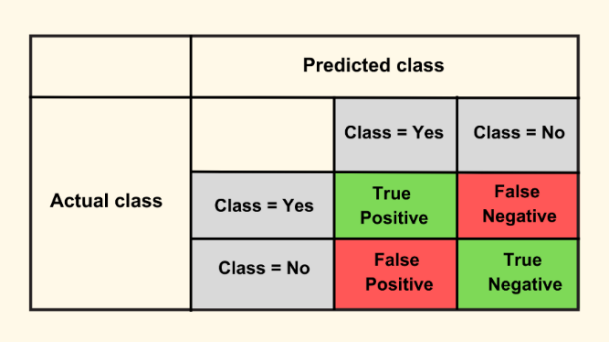

Figure 3. Comparison between Actual class and predicted class

• Precision: This measure represents the percentage of accounts that have been correctly classified as fake. The equation

expresses and conveys the precise

correctness of the model.

$$precision = T\_P/T\_P + F\_P \quad (1)$$

• Recall: What counts is the proportion of true positives that the testing model correctly categorizes. The equation provides it,which is also referred to as sensitivity.

$$recall = T\_P/T\_P + F\_N \quad (2)$$

• F-Score: It is the average of precision and recall. It comes from:

$$F1 = 2 * precision * recall \,/\, precision + recall \quad (3)$$

• Weighted Accuracy ($W - acc$): It is provided by: and indicates how frequently the test model yields the desired results.

$$w\_acc = \lambda T\_N + T\_P \,/\, \lambda\,(T\_N + F\_P) + F\_N + T\_P \quad (4)$$

Since false positives are more expensive than false negatives, λ weight is used to penalize them.

$$T\_P = True - positives;$$
$$T\_N = True - negatives;$$
$$F\_P = False - positives;$$
$$F\_N = False - negatives$$

Our LR-GD model is able to classify the accounts into two types: fake and genuine. LR and Decision Tree are also able to bifurcatethe accounts as genuine and fake with 89% and 91.6% accuracy, whereas SVM provides up to 80.8% accuracy. When compared with SVM, logistic regression and decision trees perform better in finding precision and F-score, with the highest of 89%.

6. Experiment Results and Discussions

We used data from Instagram accounts to evaluate our strategy. We selected 7,500 accounts that had been created over a six-monthperiod to use as training data, of which 55% had been flagged as spam or false. For the purpose of training our classifier, we combined Basic level labels into cluster-level labels. We used logistic regression and Gradient descent optimizer, on a comparisonwith other models we tested with support vector machine (SVM), Naive Bayes, Logistic Regression without any optimizer and Decision Tree. With a more recent dataset, we used an 80-20 split in-sample test to assess the performance of the classifiers. The weight is updated after one forward pass with one batch. That is, the weights are updated more regularly. Given those models are trained on historical data and used to simulate real-world performance, the latter test is a better representation of a performance in practice. We calculated the accuracy and recall with a precision of 95% to assess the effectiveness of the classifiers.

On experiencing the results, we came to know that the logistic regression with Gradient descent Algorithm method produced the perfect and accurate results in all measures. As compared, we all other above-mentioned classifiers in Table1, the LR-GD model generated an accuracy of92.70% and a recall of 89% with F1-Score of

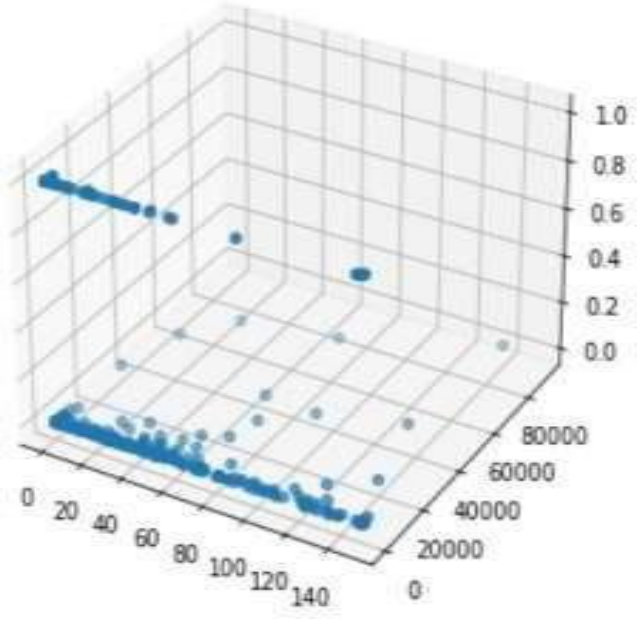

Figure 4. Disparity of Accounts Based on number of followers and posts

92%. when evaluated on testing data, this model once again surpassed all other classifiers. From other references here we are decreasing the loss and cost function of our data set this is a new kind of method. We came to know that the approach to detect fake accounts can be followed by even logistic regression using theGradient descent Optimization algorithm.

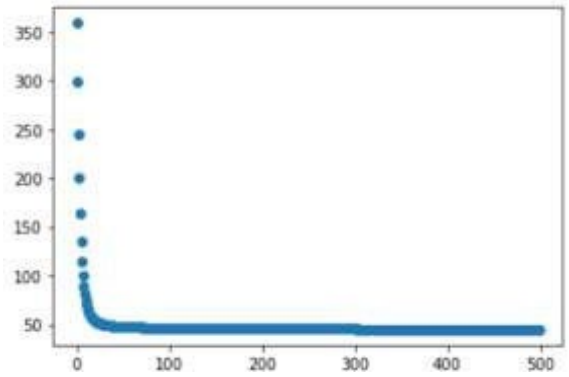

Figure 5. Gradient descent is converging correctly for 500 iterations and with learning rate = 0.5

The cost function plays a crucial role in optimizing the model's parameters. The cost function measures the difference between the predicted values and the actual labels of the training data. By minimizing the cost function, we aim to find the optimal set of parametersthat best fit the data.

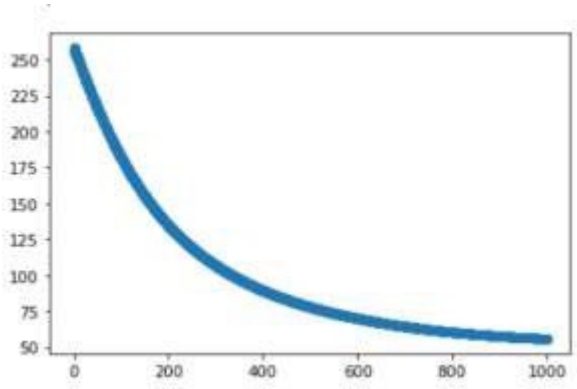

Figure 6. Cost function for alpha value 0.01

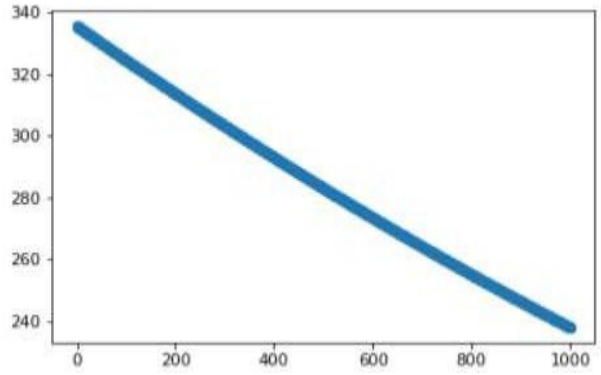

Figure 7. Cost function for alpha value 0.001

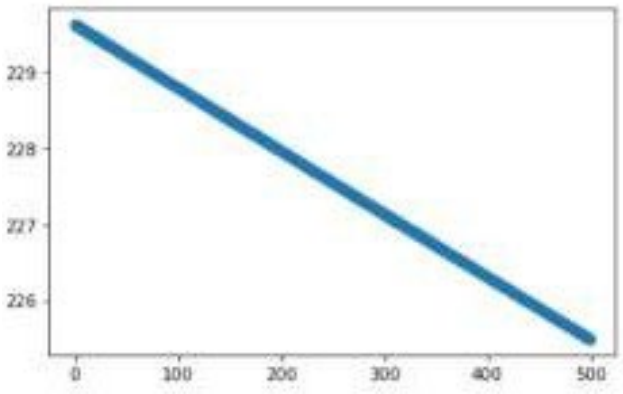

Figure 8. Cost function for alpha value 0.0001

When using different alpha values (learning rates) such as 0.01 Figure (7), 0.001 Figure (8), and 0.0001 Figure (9), it impacts the convergence and accuracy of the model. A higher learning rate like 0.01 allows for larger steps in parameter updates, which can lead to faster convergence but may also risk overshoot the optimal solution. On the other hand, smaller learning rates like 0.001 or 0.0001 take smaller steps, resulting in slower convergence but potentially better precision in reaching the global minimum. The choice of alpha depends on the dataset and the complexity of the problem. As our dataset is large and the problem is complex, a smaller learningrate may be preferable to ensure accurate convergence. However, if the dataset is smaller or the problem is relatively simpler, a higherlearning rate can help speed up the convergence process.

It is essential to experiment with different alpha values and evaluate the corresponding cost function values during training as results are shown above. The goal is to select the alpha value that achieves the lowest cost function while maintaining stable convergence. Fine-tuning the learning rate is crucial for achieving optimal performance in fake profile detection using logistic regression and the gradient descent algorithm.

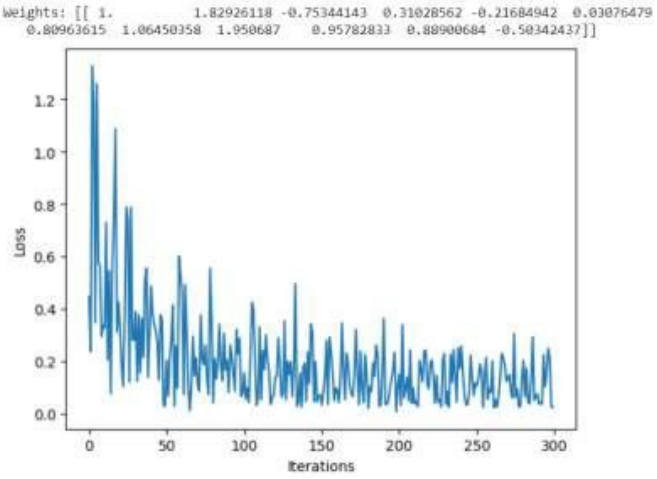

Figure 9. Trained Model Graph after Updating the Weights

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h\theta(x), y)$$

$$Cost(h\theta(X), y) = -y\log\big(h\theta(X)\big) - (1-y)\log(1 - h\theta(X))$$

While not converged {

$$\theta_j^{new} = \theta_j^{old} - \alpha\frac{1}{m}\sum_{i=1}^{m}\big(h\theta\big(x^{(i)}\big) - y^{(i)}\big)x_j^{(i)} \quad \text{for j = 0, 1...., n}$$

}

Figure 10. Logistic Regression Cost Function and Gradient Descent Update Formula

7. Limitations and Future Work

This section examines the drawbacks of our existing research and offers concepts for additional investigation. First, due to concerns about personal privacy, data collecting for OSN studies has become more challenging. As previously indicated, academics may be able to overcome this issue by seeking collaboration with OSN businesses. Second, the effectiveness of using machine learning techniques to feature the dataset was confirmed by the current investigation. This method, which has been utilized in earlier studies, would have produced more accurate comparisons. However, since the datasets came from many sources and the various methodologies make use of various assumptions, fair and equal comparisons are exceedingly challenging. In our study, suspicious accounts were not immediately identified; instead, we concentrated on creating a large enough sample size of both phony and legitimate accounts to assure reliable data. The datasets used in our paper were long-term scrapes from Sina Weibo. These accounts span times are substantially longer than their active times. The time frame is sufficient to categorize both hidden suspicious accounts and regular accounts. The threshold of time for creating the account in real apps is connected to the active time frame in some OSNs. The time barrier might have a negligible effect on a real-world platform's ability to detect and combat concealed suspicious accounts.

The future scope of using logistic regression and the gradient descent algorithm for fake profile detection holds great potential for further advancements and applications. Some key areas of future exploration and development include:

Integrating deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), holds immense potential for enhancing the performance of fake profile detection. By incorporating deep learning models into the existing framework, we can leverage their ability to learn intricate representations and capture complex relationships within the profile data. This integration enables the detection system to gain a deeper understanding of the underlying patterns and features that distinguish real profiles from fake ones, consequently leading to more accurate and robust classification.

Convolutional neural networks (CNNs) are particularly effective in analyzing visual data, making them well-suited for extracting meaningful information from profile images. By employing CNNs, the detection system can automatically learn hierarchical representations of image features, such as facial characteristics, image quality, or inconsistencies, which are indicative of fake profiles. The learned representations enable the model to discern subtle visual cues and identify manipulated or synthetic images commonly used in fraudulent profiles. Additionally, recurrent neural networks (RNNs) are instrumental in modelling sequential and temporal dependencies within profile data. RNNs can capture the sequential nature of textual content, such as profile descriptions or posts, and effectively analyze patterns and linguistic cues that differentiate genuine profiles from fake ones. By considering the contextual information and dependencies within the text, the model becomes more adept at detecting anomalies, grammatical inconsistencies, or suspicious patterns that are prevalent in fake profiles. The integration

of deep learning techniques also facilitates the extraction of high-level representations that encapsulate both visual and textual information, by combining the outputs of CNNs and RNNs, or employing more advanced architectures like multi-modal networks, the detection system can exploit the complementary nature of different data modalities. This fusion of information allows for a more comprehensive analysis of profile data, capturing nuances andcorrelations that might be missed by using each modality independently.

Furthermore, the integration of deep learning techniques supports end-to-end learning, where the model learns directly from raw datawithout the need for manual feature engineering. This enables the detection system to automatically discover and adapt to relevant features, reducing the dependence on handcrafted features and potentially uncovering previously unknown patterns that contribute tofake profile identification. As deep learning models continue to evolve and improve, the integration of these techniques into the fakeprofile detection framework promises to unlock new possibilities for more accurate and robust classification. The ability to learn complex representations and capture intricate relationships inherent in deep learning models aligns well with the nuanced nature of fake profile detection, leading to heightened detection performance and a stronger defense against fraudulent activities in online platforms and social networks.

8. Conclusion

In OSNs, suspicious accounts have advanced significantly in recent years and can now successfully blend in with legitimate accounts. In this work, we started by extracting features from the hidden suspect accounts and training the model. the feature analysis makes it easier to find hidden suspicious accounts, which are frequently missed. We found these features and assessed them in a detection method to compare the effectiveness of fake and real accounts. The accuracy and FP rates (92.7% and 0.5%,respectively) were excellent. Finally, we contrasted the various classifiers developed using features from single-account features and features from the Logistic regression with gradient descent. This model's efficiency is demonstrated by a reduction in the cost function, and its accuracy in forecasting outcomes is further supported by feature assessment ranking. Overall, the evaluation's findings demonstrated the value and necessity of employing LR-GD to identify concealed, sporadic suspect accounts.

References

[1] Devakunchari Ramalingam∗ , Valliyammai Chinnaiah (2021) Fake profile detection techniques in large-scale online socialnetworks: A comprehensive review

[2] Xiang Zhu, Yuanping Nie, Songchand Jin, Aiping Li, and Yan jia, "Spammer Detection on Online Social Networks Based onLogistic Regression," Springer International Publishing Switzerland 2015 X. Xiao and Z. Zhang (Eds.): WAIM 2015, LNCS 9391, pp. 29–40, 2015. DOI: 10.1007/978-3-319-23531- 8 3

[3] Dong Yuan, Yuanli Miao, Neil Zhenqiang Gong, Zheng Yang, Qi Li1, Dawn Song, Qian Wang, Xiao Liang "DetectingFake Account in Online Social Networks at the time of registrations," Association for Computing Machinery. ACM ISBN 978-1-4503-6747-9/19/11. . . $15.00 https://doi.org/10.1145/3319535.336319 8

[4] Cao Xiao, David Mandell Freeman and Theodore Hwa, "Detecting clusters of fake accounts in online social networks," 2015 ACM. ISBN 978-1-4503-3826- 4/15/10 ...$15.00. DOI: http://dx.doi.org/10.1145/2808769.2808779

[5] Ajesh F, Aswathy S U , Felix M Philip and Jeyakrishnan V, "A Hybrid Method for Fake Profile Detection in Social Network Using Artificial Intelligence", SecurityIssues and Privacy Concerns in Industry 4.0 Applications, (89–112) © 2021Scrivener Publishing LLC

[6] Monika Singh, Divya Bansal and Sanjeev Sofat, "Who is Who on Twitter–Spammer, Fake or Compromised Account? A Tool toReveal True

Identity in Real-Time, Cybernetics and Systems", DOI: 10.1080/01969722.2017.1412866

[7] Mohammadreza Mohammadrezaei, Mohammad Ebrahim Shiri and Amir Masoud Rahmani, "Identification Fake Account onSocial Network Based on Graph Analysis and Classification Algorithms", Hindawi Security and Communication Networks, https://doi.org/10.1155/2018/5923156

[8] Fatih Cagatay Akyon and M. Esat Kalfaoglu, "Instagram Fake and Automated Account Detection," 978-1-7281-2868-9/19/$ 31.0 2019 IEEE

[9] Mazhar Javed Awan, Muhammad Asad Khan, Zain Khalid Ansari, Awais Yasin, Hafiz Muhammad Faisal Shehzad, " Fake profile recognition using big data analytics in social media platforms," https://doi.org/10.1504/IJCAT.2022.124942

[10] B. Prabhu Kavin,Sagar Karki,S. Hemalatha,Deepmala Singh,R. Vijayalakshmi,M. Thangamani,Sulaima Lebbe Abdul Haleem,Deepa Jose,Vineet Tirth,Pravin R. Kshirsagar,and Amsalu Gosu Adigo," Machine Learning-Based Secure Data Acquisition for Fake Account Detection in Future Mobile Communication Networks," https://doi.org/10.1155/2022/6356152

[11] Faouzia Benabbou, Hanane Boukhouima, Nawal Sael "Fake accounts detection system based on bidirectional gated recurrentunit neural network," ISSN:2088-8708, DOI: 10.11591/ijece.v12i3.pp3129-3137

[12] N. Kanagavalli1, S. Baghavathi Priya, " Social Networks Fake Account and Fake News Identification with Reliable Deep Learning ," DOI:10.32604/iasc.2022.022720

[13] Aditya, B.L.V.S., Rajaram, G., Hole, S.R., Mohanty, S.N. (2023). F2PMSMD: Design of a Fusion Model to Identify Fake Profiles from Multimodal Social Media Datasets. In: Nandan Mohanty, S., Garcia Diaz, V., Satish Kumar, G.A.E. (eds) Intelligent Systems and Machine Learning. ICISML 2022. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 471. Springer, Cham. https://doi.org/10.1007/978-3-031-35081-8_2

[14] B. L. V. S. Aditya and S. N. Mohanty, "Heterogenous Social Media Analysis for Efficient Deep Learning Fake-Profile Identification," in IEEE Access, vol. 11, pp. 99339-99351, 2023, doi: 10.1109/ACCESS.2023.3313169.