Global Scientific JOURNALS

GSJ: Volume 8, Issue 6, June 2020, Online: ISSN 2320-9186
www.globalscientificjournal.com

Social Media Fake Account Detection for Amharic Language by using Machine Learning

Kedir Lemma Arega Shewa, Ethiopia
Email: kedirnaw1999@gmail.com
School of Technology and Informatics, Ambo University

Abstract

A social networking service serves as a platform to build social networks or social relations among people who, share interests, activities, backgrounds, or real life connections. A social network service is generally offered to participants who registers to this site with their unique representation (often a profile) and one's social links. Most social network services are web-based and provide means for users to interact over the Internet. [1]. Online social networking sites became an important means in our daily life. Millions of users register and share personal information with others. Because of the fast expansion of social networks, public may exploit them for unprincipled and illegitimate activities. As a result of this, privacy threats and disclosing personal information have become the most important issues to the users of social networking sites. The intent of creating fake profiles have become an adversary effect and difficult to detect such identities/malicious content without appropriate research. The current research that have been developed for detecting malicious content, primarily considered the characteristics of user profile. Most of the existing techniques lack comprehensive evaluation. In this work we propose new model using machine learning and NLP (Natural Language Processing) techniques to enhance the accuracy rate in detecting the fake identities in online social networks. We would like to apply this approach to Facebook by extracting the features like Time, date of publication, language, and geo position. [2]

Key words: *Amharic, Classification, Detection, Fake account, Machine learning, NLP, Social media,*

1. Introduction

1.1. Background

Social media currently provide localization, which allows the user to use different world languages on their sites. One of these languages is Amharic, Amharic languages are one of wildly spoken language and working language of the federal government of Ethiopia. The language is written left-to-right and has its unique script, which lacks capitalization and in total 275 characters, mainly consonant-vowel pairs. [3] It is the second most spoken Semitic language in the world (after Arabic) and closely related to Tigrinya. It is probably the second largest language in Ethiopia (after Oromo, a Cushitic language) and possibly one of the five largest languages on the African continent. Despite the relatively large number of speakers, Amharic is still a language for which very few computational linguistic resources have been developed for the language. [3]

Online Social Networks are most popular through which information can be exchanged through the world. Social Networks being the center of attraction for many applications and they incorporate a range of new information and communication tools to the user community. A Social Network is best viewed as a graphical structure with nodes and edges depicting the users and their interaction activities respectively. The nodes and edges in a Social Network graph can be labeled or unlabeled depending upon the structure of the network being used. Because of the great reputation of social intelligence, social networking sites such as Facebook, YouTube, Twitter, LinkedIn, Pinterest, Google +, Tumblr and Instagram have become the preferred means of communication and information sharing tools amongst a diverse set of users including individuals and companies. The users of the social networks will play a vital role and they are completely responsible for the contents being exchanged in the networks. Users share information by interesting websites, videos and files. People share confidential data through the set-up of great faith and others have the same faith in the data shared. The rush of online social networks' reputation and the accessibility of huge amount of data enable them simple objective to the opponents. These objectives mainly include stealing individual user's details without seeking any permission. One of the main problems in social media is the spammers as they can use their accounts for different targets. One of these targets is spreading rumors which may affect a determined business or even the society in a large scale. According to the importance of the effect of social media to the society, in this research, [4] aim to detect the fake profile accounts from Twitter online social network to prevent the spreading of fake news, advertisements and fake followers.

The attempt for the encroachment of a legitimate user profile through fake identities is considered as the mostly practiced technique. As the expansion of greater security in online social networking sites it turned to be very hard to encroach into online social networks. As a result of this, antagonists create false identities to gain access to other profiles. [2] In 2019, Facebook took down on average close to 2 billion fake accounts per quarter. Fraudsters use these fake accounts to spread spam, phishing links, or malware. It's a lucrative business that can be devastating for any innocent users that it snares. Facebook is now releasing details about the machine-learning system it uses to tackle this challenge. The tech giant distinguishes between two types of fake accounts. First, there are "user-misclassified accounts," personal profiles for businesses or pets that are meant to be Pages. These are relatively straightforward to deal with—they just get converted to Pages. "Violating accounts," on the other hand, are more serious. These are personal profiles that engage in scamming and spamming or otherwise violate the platform's terms of service. Violating accounts need to be remove as quickly as possible without casting net and snagging real accounts as well. [5] The main objective of any Social Networking Site is to target different user segments. The best thing about Facebook is the ability to find old friends, but YouTube provides a platform for people to connect, inform, and inspire others across the world by video sharing. According to ETV News (Ethiopian Television) report in June 5, 2020 more than 5 million Birr (money) were fraud by fake account user in social media. The following figure shows how the fake account is a serious problem. [6] Available on: https://www.youtube.com/watch?v=e9s3B4dZJus

Figure 1 fake account and fraud.

1.1.1. The Amharic Character Representation

Amharic utilizes Geez characters; the characters trace back to 4th century A.D. The first forms of the Geez script included only consonants, while the subsequent variants of the characters represent phoneme pairs of consonant-vowel. Like Geez, Amharic writing uses characters formed by a consonant-vowel combination. In Amharic, seven vowels are used, each in seven distinct forms that reflect the seven vowel sounds they are ሀ ፡ ሁ ፡ ሂ ፡ ሃ ፡ ሄ ፡ ህ ፡ ሆ. There are 33 basic characters with seven forms representing a consonant and a vowel at the same time, which makes the Amharic script pronounced in the syllable. The first order is the basic form, and there are 33 basic forms with six derivations for each giving 231 characters [3] Now a days use of internet is increased. with the use of internet, the term social media networks become popular. Everyone who use internet is well-known about social media networks. Social media network is collection of many social networking websites. Social networking is platform, where a user of social network can express their point of view towards anything. [7]

1.1.2. Amharic Punctuation

The Amharic language has around ten punctuation marks in but few of them used in a computer system. Also, most of them are sentence separator marks. Punctuation mark such as ፡ (hulet neteb)/ (word separator or space), ። (Arat Neteb)/ (full stop (period)), ፣ (Netela Serez)/(comma), and ፤ (Dereb Serez)/(semicolon). [3]
Online social networks (OSNs), such as Facebook, Twitter, RenRen, LinkedIn, Google+, and Tuenti, have become increasingly popular over last few years. People use OSNs to keep in touch with each others, share news, organize events, and even run their own e-business. [8]

1.2. Principal Component Analysis

PCA is applied to reduce the dimensionality of the dataset. In this proposed work PCA plays an important position by giving the great endorsement to make decisions on which profile features to be used. Principal Component Analysis (PCA) is the simplest and robust dimensionality reduction technique ever seen. In this paper we have selected a mathematical model called variance maximization for drawing PCA results. According to this model “first principal component has the highest projection variance which is the direction in feature space along. And the second component defines the direction which has highest projection variance among all the other orthogonal direction to the first component”. While calculating the score on profile features both false and real accounts to be measured [9]

1.3. Related Work

Different researches have been presented to detect fake accounts with different approaches in this study, they have presented a classification method for detecting the fake accounts on Twitter. They have preprocessed the dataset using a supervised discretization technique named Entropy Minimization Discretization (EMD) on numerical features and analyzed the results of the Naïve Bayes algorithm. [4]. Inspired by the importance of detecting fake accounts, researchers have recently started to investigate efficient fake accounts detection mechanisms. Most detection mechanisms attempt to predict and classify user accounts as real or fake (malicious, Sybil) by analyzing user level activities or graph-level structures. There are several data mining methodologies [4] and approaches that help detecting fake accounts that are described in the following sub-sections. [7] In this section, we woud demonstrate some of the works that have been presented in this area. Reference [1]has reached an accuracy 80% the performance were evaluated using the supervised machine learning algorithms and the highest accuracy were obtained and the maximum percentage of skin exposed were calculated from the images collected from the fake accounts. However, in my research. [10]Neural network algorithm is used to evaluate the proposed feature set and compare it against the state-of-the-art feature sets in detecting fraud. The feature set considers the user’s social interaction on the Yelp platform to determine if the user is committing fraud. The neural network algorithm helps in comparing the feature set with other feature sets used to detect fraud. Any attempt to find the characteristics that lead to fraud has a prerequisite to be good enough to detect fraud as well. However, [11] OSNs suffer from abuse in the form of the creation of fake accounts, which do not correspond to real humans. Fakes can introduce spam, manipulate online rating, or exploit knowledge extracted from the network. OSN operators currently expend significant resources to detect, manually verify, and shut down fake accounts. [12]Information is spread across social networks quickly. However at the same time social media networks become susceptible to different types of unwanted spammer actions. As part of their work, they propose a mechanism to detect spammers in facebook social network. Their work is based on number of features at content level and user level. Use [13]classification algorithms in machine learning to detect fake ac-

counts. The process of finding a fake account mainly depends on factors such as engagement rate and artificial activity. and Decision trees are made seeing the success rate i.e., in their case taking the value which contains more fake accounts. Following Table show works done by different Peoples in this area. [1], [4], [14], [9], [10], [15], [12], [8]

Author and year	Title	Feature extraction	Method and Accuracy
M. Smruthi, N. Harini (2019)	A Hybrid Scheme for Detecting Fake Accounts in Facebook	machine learning and NLP (Natural Language Processing) techniques	Time, date of publication, language, and geoposition
Buket Ersahin1, Ozlem Aktas1 , Deniz Kilinç2, Ceyhun Akyol2 (2017)	Twitter Fake Account Detection	supervised discretization technique named Entropy Minimization Discretization (EMD	85.55%
Mohammadreza Mohammadrezaei ,1 Mohammad Ebrahim Shiri ,1,2 and AmirMasoud Rahmani1,3,4 (2018)	Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms	Graph Analysis and Classification Algorithms	75%
Srinivas Rao Pulluri1, Jayadev Gyani2, Narsimha Gugulothu3 (2017)	A Comprehensive Model for Detecting Fake Profiles in Online Social Networks	machine learning and NLP (Natural Language Processing) techniques	Time, date of publication, language, and geoposition
Kunal Goswami, Younghee Park* and Chungsik Song (2017)	Impact of reviewer social interaction on online consumer review fraud detection	machine learning techniques	F-score of 75.4 % for burst reviews, and 68.7 % for all reviews.
Michael Crawford*, Taghi M. Khoshgoftaar, Joseph D. Prusa, Aaron N. Richter and Hamzah Al Najada (2017)	Survey of review spam detection using machine learning techniques	machine learning techniques	65 % accuracy
K Subba Reddy, Dr E Srinivasa Reddy (2017)	An Efficient Methodology to Detect Spam in Social Networking Sites	Naïve Bayes and Decision Tree algorithms	The integrated algorithm classifies an account as spammer or non spammer with an overall accuracy of 90.5%.
Sarah Khaled, Hoda M. O. Mokhtar, Neamat El-Tazi (2018)	Detecting Fake Accounts on Social Media	classification	Roughly 70% of spammers and 96% of non-spammers were effectively characterized in their outcome.

Table 1. Summary of related work

1.4. Proposed Algorithm

This section presents the proposed methods of predicting fake twitter accounts. Proposed methods are divided into two main parts: feature reduction, and data classification aiming to develop a new technique that achieves a high classification accuracy results in a reasonable time. [7]

1.4.1. Data Pre-Processing

The ”MIB” dataset feature vectors are presented in two types:

- Categorical features e.g. language, profile-side bare color, tweets.
- Numerical features e.g. friends-count, followers count, default-profile, profile-use-background image. [7]

1.4.2. Building Dataset

The objectives of this study are to fake account detection in social media and it needs to build a new Amharic dataset. This new dataset needed because there is no published or annotated dataset for this purpose. The process of building the dataset for Amharic fake account consists of the following main steps,

1. Gathering the Amharic post and comment textual data from public Facebook pages

2. Preparing, filtering, or consolidating gathered data into one file dataset. And

3. Annotating the dataset.

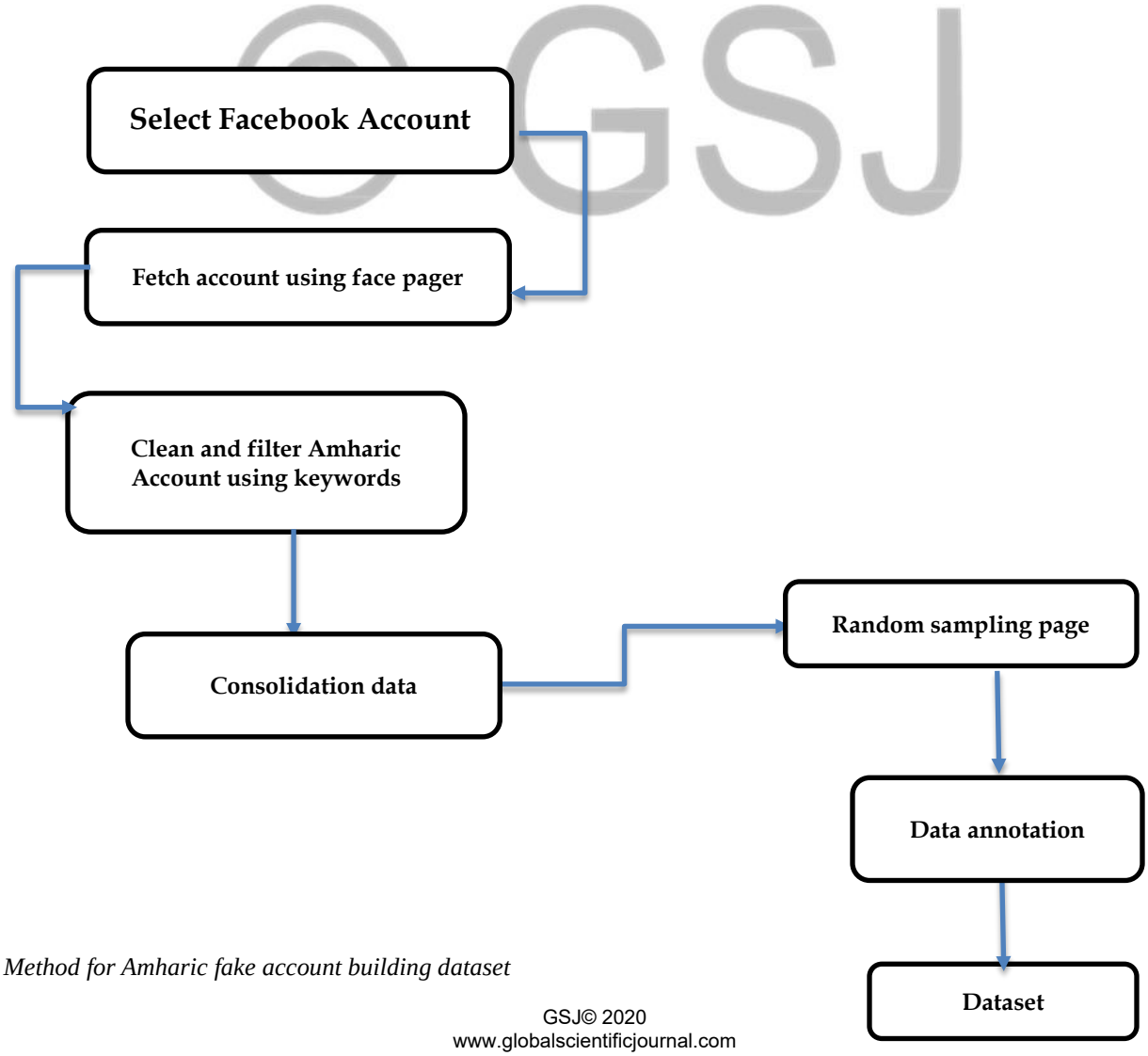


Figure2 Method for Amharic fake account building dataset

1.4.3. Feature Reduction

In feature reduction phase, four data reduction techniques were applied to guide the process of deciding the most promising feature patterns to be used in the mining process [7]

- PCA
- Spear mans Rank-Order Correlation
- Wrapper Feature Selection using SVM
- Multiple Linear Regression

1.4.4. Selection of tool

In this study, a number of tools are used in order to come up with the solution for the problem that is going to be addressed. Different tools are used for the development of the proposed detection system. Java programming language used for the development of the detection model. Java support platform independence and it is suitable for encoding Unicode. The development tools will be used for implementing Python 3.3.3 depending on the situations as they necessity.

1.4.5. Experiment

Experiments are performed to evaluate the performance of the developed system as the following flow chart

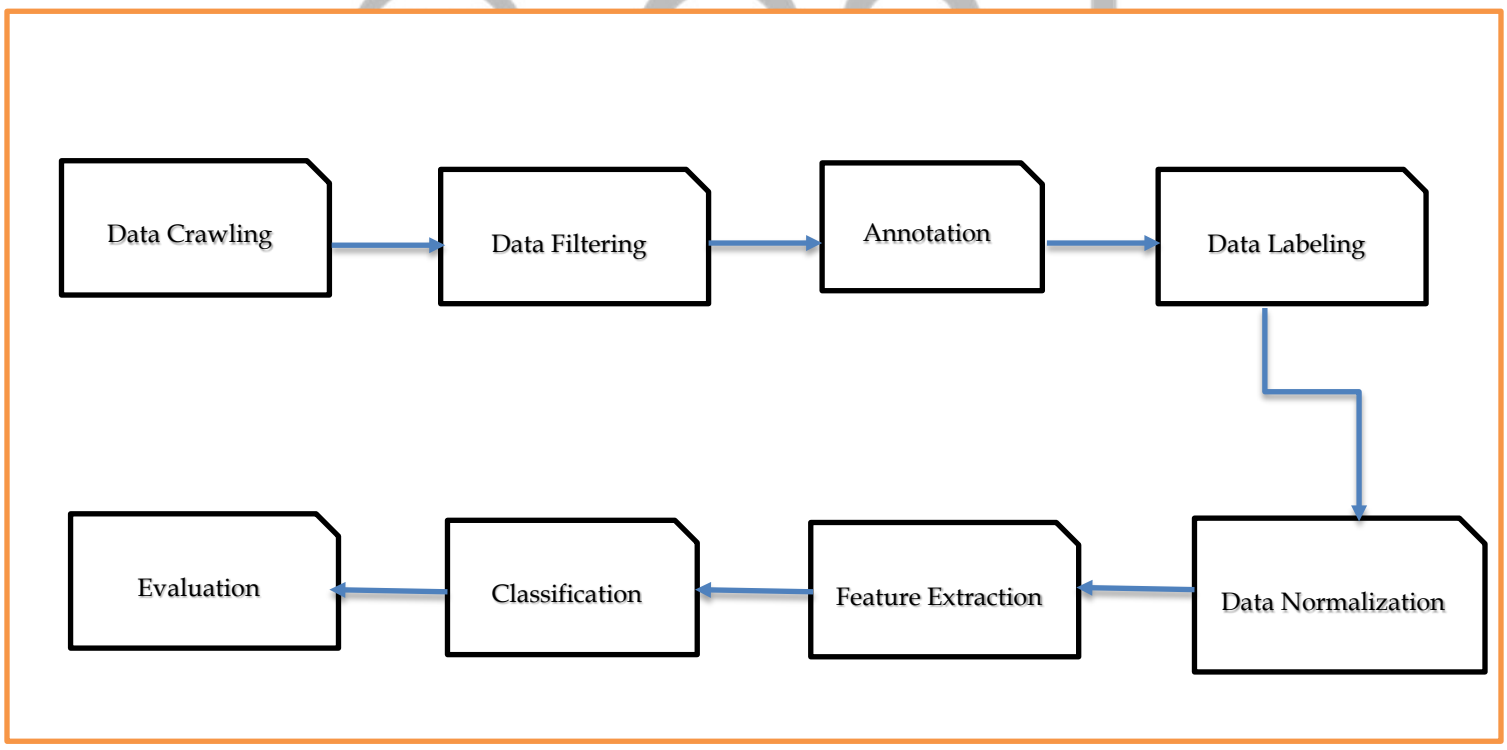


Figure 3 The proposal experiment flowchart

1.5. Performance and Evaluation

In this section the results and findings of this work would be explained and evaluated. Initially, three different classification algorithms have been trained and tested using divergent four feature sets. Neural network classifi-

cation algorithm and SVM classification algorithm were used as the principles mining techniques in many social network researches, so they have been applied on the feature sets mentioned in Feature Reduction and compared with the proposed SVN-NN algorithm. [7]

```
|
Results graph method
Number of loops: 55
              precision    recall  f1-score   support

         0.0       0.88      0.55      0.67       251
```

Figure 4 Performance evaluation using Graph Method

```
svm train accuracy 0.9281200631911533
svm test accuracy 0.8958990536277602
SVM algorithm results
              precision    recall  f1-score   support

         0.0       0.91      0.97      0.94       251
         1.0       0.84      0.62      0.71        66

   micro avg       0.90      0.90      0.90       317
   macro avg       0.87      0.79      0.82       317
weighted avg       0.89      0.90      0.89       317

229 1037
```

Figure 5 Performance evaluation using SVM

```
Method results combining both characteristic and graph
Number of loops: 44
              precision    recall  f1-score   support

         0.0       0.96      0.98      0.97       251
         1.0       0.90      0.85      0.88        66

   micro avg       0.95      0.95      0.95       317
   macro avg       0.93      0.91      0.92       317
weighted avg       0.95      0.95      0.95       317
```

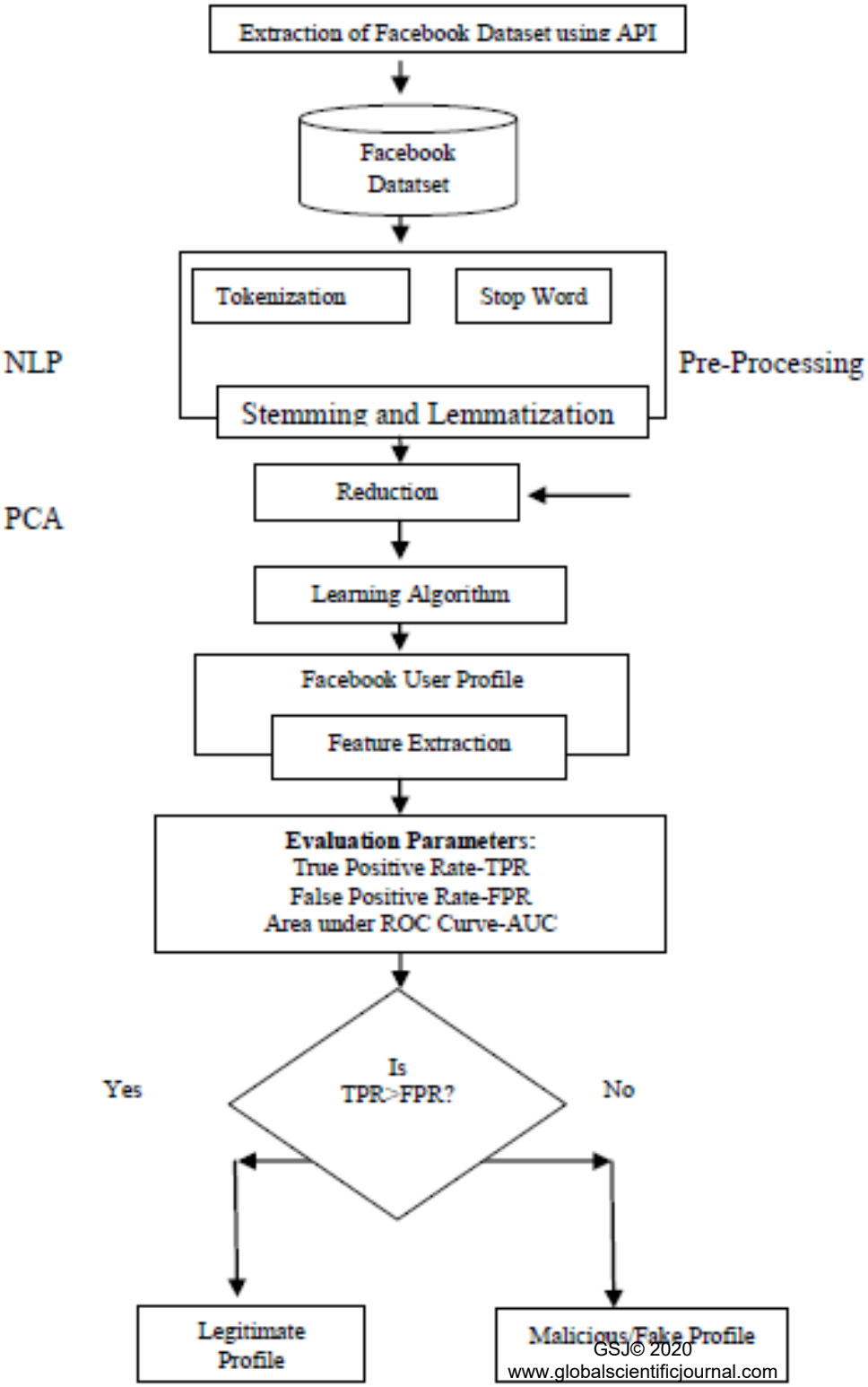
Figure 6 Both Graph and SVM performance evaluation

1.5.1. **Neural Networks**

Currently, there are many neural network algorithms used to train models and predict results based on the previously trained models. Feed-forward back propagation algorithm has been selected as the base algorithm. The predicted results have been compared with the actual legitimate values (i.e. whether the account is real or fake), and the prediction accuracy was calculated as follows:

$$\%Accuracy = \frac{All\ correctly\ identified\ accounts}{Total\ number\ of\ accounts} \times 100$$

As mentioned above the feature subsets with highest accuracy was highlighted, as following: spearmans rank-order Correlation best pattern was (1000001000110110), Multiple linear Regression best pattern was (0110110111001111), Wrapper-SVM best pattern was (110111111011111). [7]



Most of the existing techniques for detecting malicious content of Facebook lack inclusive evaluation. The main objective of [2] research work is to increase the accuracy rate in identifying the fake profiles/malicious content in online social networking sites as compared to existing research. We would like to apply the proposed approach on Facebook.

Working Principle of Proposed Work

Figure 7 working principle of proposed work

1.6. Application Result

User activities related to likes, comments, and to some extent, shares on Facebook, contribute the maximum to detection of fake accounts. Therefore, this work represents a significant step towards a profile-feature based detection of fake accounts on Facebook. Many fake

users were classified as real, possibly because fake accounts mimic real user behavior to elude detection mechanisms.
Detecting and blocking fake account is important for online communities for maintaining safe environments for its real users and as a responsibility considering their impact on society. Fake account detection system will help for reduction of time, fraud and human effort to identify privacy attack on social media. The system will help to filter any fake user that makes peoples of the local population indirectly or directly participate in the violent activities across the different region of the country.

1.7. Conclusion

Fake accounts are being continuously evolving in online social media. Therefore, it is very essential to invent new methods to detect Fake profiles in online social media. So the real time Facebook dataset were required to detect the fake accounts and vulgar images in Facebook. For the detection of Fake accounts the user timeline information namely post-count, comment-count, etc. were used and for the vulgar image detection the images from the user time line and the display picture of the users were taken out. The performance were evaluated using the supervised machine learning algorithms and the highest 80%accuracy were obtained and the maximum percentage of skin exposed were calculated from the images collected from the fake accounts. For the future scope, a more complex algorithm for the skin detection can be implemented. The natural language processing techniques can be implemented to detect fake accounts more accurately. The new features will be certainly introduced by the Facebook, and these features can also be included while analyzing the fake accounts. [1]

REFERENCES

[1] N. H. . M. Smruthi, "A Hybrid Scheme for Detecting Fake Accounts in Facebook," *International Journal of Recent Technology and Engineering (IJRTE),* pp. 213-217, , February 2019.

[2] J. G. N. G. Srinivas Rao Pulluri1, "A Comprehensive Model for Detecting Fake Profiles in Online Social Networks," *International Journal of Advanced Research in Science and Engineering,* pp. 1-10, 2017.

[3] Y. K. Defar, "Hate Speech Detection for Amharic Language on Social Media Using Machine Learning Techniques," pp. 1-103, September 2019.

[4] Ö. A. D. K. C. A. Buket Ersahin1, "Twitter Fake Account Detection," *IEEE,* pp. 388-392, 2017.

[5] K. Hao, "Hao, Karen Archive Page," 4 March 2020. [Online]. Available: https://www.technologyreveiw.com.

[6] ETV, "News," Addis Ababa, 2020.

[7] S. B. S. A. Sachin Ingle1, "Detecting Fake User Accounts on," *IJARIIE-ISSN(O)-2395-4396,* pp. 927-931, 2019.

[8] H. M. O. M. N. E.-T. Sarah Khaled, "Detecting Fake Accounts on Social Media," in *IEEE International Conference on Big Data (Big Data)*, Cairo, 2018.

[9] J. G. N. G. Srinivas Rao Pulluri1, "A Comprehensive Model for Detecting Fake Profiles in Online Social Networks," *International Journal of Advanced Research in Science and Engineering,* p. 10, 2017.

[10] Y. P. a. C. S. Kunal Goswami, "Impact of reviewer social interaction," *Springer Journal of Big Data,* pp. 1-19, 2017.

[11] Q. C. †. M. S. ‡. X. Y. T. Pregueiro, "Aiding the Detection of Fake Accounts in Large Scale Social Online Services," pp. 1-14.

[12] D. E. S. R. K Subba Reddy, "An Efficient Methodology to Detect Spam," *International Journal of Computer Science and Information Security (IJCSIS),,* pp. 151-158, 2017.

[13] H. K. G. S. T. P. R. S. P. Maniraj, "Fake Account Detection using Machine Learning and Data Science," *International Journal of Innovative Technology and Exploring Engineering (IJITEE),* pp. 583-585, 2019.

[14] 1. M. E. S. ,. Mohammadreza Mohammadrezaei, "Identifying Fake Accounts on Social Networks Based on," *WILEY HINDAWI,* pp. 1-9, August 2018.

[15] T. M. K. J. D. P. A. N. R. a. H. A. N. Michael Crawford*, "Survey of review spam detection using machine

learning techniques," *Springer Journal of Big Data,* pp. 1-24, 2015.

[16] B. G. Erena., "Orormo Language (Afaan Oromoo)," [Online]. Available: https://scholar.harvard.edu/erena/oromo-language-afaan-oromoo.

[17] L. Guta, "Social network hate speech detection for afaan oromoo language," p. 8, 11 June 2019.

[18] C. L. P. a. N. Solomom, "Social media and journalism i Ethiopia," FOJO MEDIA INSTITUTE , Linnaeus University Stockholm, 2019.

[19] www.facebook.com, "Fake account," (MAU) on Facebook , 2019.