IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Issue 4,June 2022

Detection of Fake Profiles on Social Networks using Machine Learning ANN & SVM Algorithms

[1]Dr.D.Kavitha
Sr.Assistant Professor
Department of IT
P V P Siddhartha Institute of Technology, Kanuru, Vijayawada, Krishna district - 520007
Mail id- kavitha.d@pvpsiddhartha.ac.in

[2]K .Sri Vijaya
Assistant Professor
Department of IT
P V P Siddhartha Institute of Technology, Kanuru, Vijayawada, Krishna district – 520007
srivijayak@gmail.com

DOI: 10.48047/IJFANS/ISSUE4/002

Abstract
Social Networks plays an important role for internet users to carry out their daily activities like content sharing, news reading, posting messages, product reviews and discussing events etc. At the same time, various kinds of spammers are also equally attracted towards these social media. These cyber criminals including sexual predators, online fraudsters, advertising campaigners and trollers etc. These guys are creating fake profiles to spread their content and carry out for scams. All these malicious identities are very harmful for both the users as well as the service providers. From the social media service providers identify those accounts and check it is genuine or fake. In this Paper we proposed many classifications algorithm like support vector machine algorithm and neural network. These algorithms help to detect the fake profiles on social media.

Keywords: *Artificial Intelligence, Machine Learning, Social Networks, ANN, SVM.*

Introduction

In the present generation, everyone in society has become associated with the social media. These social media have made a drastic change in the way we pursue our social life. In this Paper using Artificial Neural Networks we will identify whether given account details are from genuine or fake users. ANN with SVM algorithm will be trained with all previous users fake and genuine account data and then if we gave new test data then that ANN train model will be implemented on new test data. It will identify whether given new account details are whether fake or genuine.

- In these days social media has dangerous impact affecting mental health of people.
- Many people are getting proned to fake profiling.
- The community of concern to us here is fake Accounts and our problem can be said to be a classification or a clustering problem.
- To avoid such problems fake profiling detection can provide a better solution.

Each profile (or account) in a social network contains many information like gender, no. of friends, no. of comments, education, work, etc a number of this information is private and some are public. Since private information isn't accessible So, we've used only the knowledge that's public to determine the fake profiles within the social network However, if our proposed scheme is employed by the social networking companies itself then they will use the private information of the profiles for detection without violating any privacy issues. we've considered this information as features of a profile for the classification of fake and real profiles. The steps that we've followed for the identification of fake profiles.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Issue 4,June 2022

Scope of the Paper is to identify the profiles accuracy whether the given profile is genuine or fake.

Related Work

Accounts in online social media have heaps of input data like name, sexual orientation, companions, devotees, preferences, area numbers. Half part of this input data is both of public and private. We have to use input that are public to know profiles which are phony for interpersonal organization as data from private is unavailable[1][2].

A confusion matrix is a summary of prediction outcomes on a classification problem. The number of accurate and incorrect predictions are summarized with depend values and damaged down by each elegance. That is the key to the confusion matrix[3][4]. The confusion matrix suggests the methods in which your classification model is confused while it makes predictions[5].

Methodology

The Application Domain of the subsequent Paper was Community detection[6]. Community detection is vital to understand the structure of complex networks and ultimately extract information from them. During this Paper, a framework is used through which a fake profile is detected using machine learning algorithm in order that the social lifetime of people become secured[6][7].

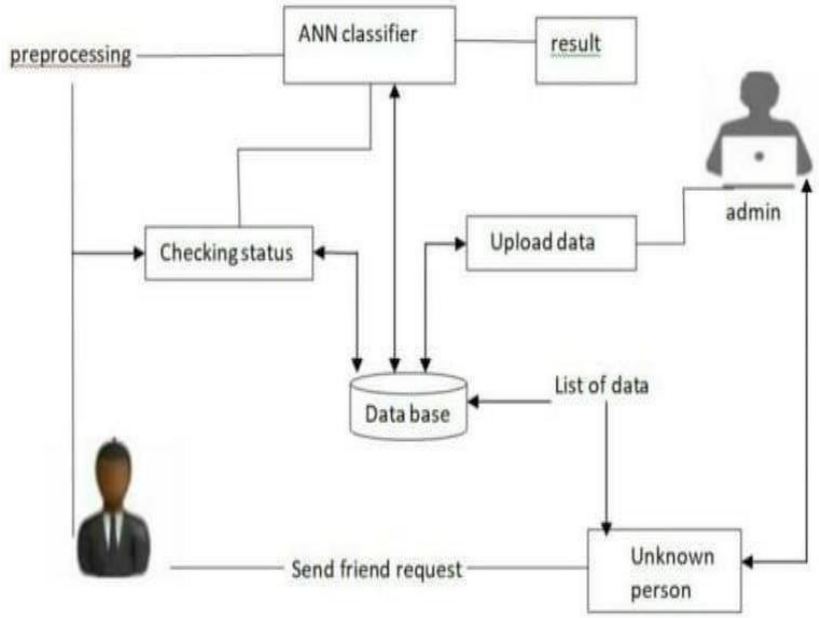

Figure 1: Architectural Diagram

Fig 1 depicts the system that is used to abstract the overall outline of the software system and the relationships, constraints, and boundaries between components. It is an important tool as it provides an overall view of the physical deployment of Fig 2 software system and its evolution roadmap.

In Fig 1 a user sends a friend request to an unknown person and the entire data of that unknown person is stored in database and ANN classifier preprocess and check the status of that user from the database and admin uploads the data in the database and then produces the result using ANN classifier[8][9].

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Issue 4,June 2022

Classification starts from the choice of profile that must be classified. Once the profile is chosen, the useful features are extracted for the aim of classification. The extracted features are then fed to trained classifier. The results of classification algorithm are then verified, and feedback is fed back to the classifier[10].

Procedure to detect Fake Profile using Neural Networks.

For the detection of fake profiles in online social networking site using Neural network methodology, following steps have been taken into consideration:

Step 1: Firstly, we have imported libraries of sys, csv, os, date time, math, NumPy, pandas, and matlab.

Step 2: After this, gender detection libraries are being loaded to compute the information about the gender. For validating the data and preprocessing, sklearn libraries have been integrated to plot the matrix. The evaluation metric provides the information about different variables of confusion matrix. For the evaluation of classifier, area under cover and accuracy have been used.

Step 3: Then, import the Pybrain library for training the datasets. It is freely available open-sourced library for machine learning algorithms. Different utility tools get implemented along with this library.

Step 4: Next is to read the dataset by defining a function name read_datasets(). The CSV or comma separated value files get used for this. We need to set the default for datasets to read. After combining the fake and authentic user, length of users needs to be found.

Step 5: Afterward, another function is defined to predict the gender of the person through the given the name. First name of the person gets declassified into parts for computing the model. Further, different features associated with that will be integrated along with status counts, and followers counts and so on.

Step 6: Next, plotting of confusion matrix begins which integrate the plot as per the fake and genuine profile accounts.

Step 7: Function for defining the ROC curve has been implementing for further computation.

Step 8: Using the neural network, function is declared to train the dataset. For this, read_datasets() has been used.

Step 9: After reading the data, the output will look like this.

Step 10: Extracting the features of the training datasets.

Step 11: The graph of confusion matrix without normalization has been shown.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Issue 4,June 2022

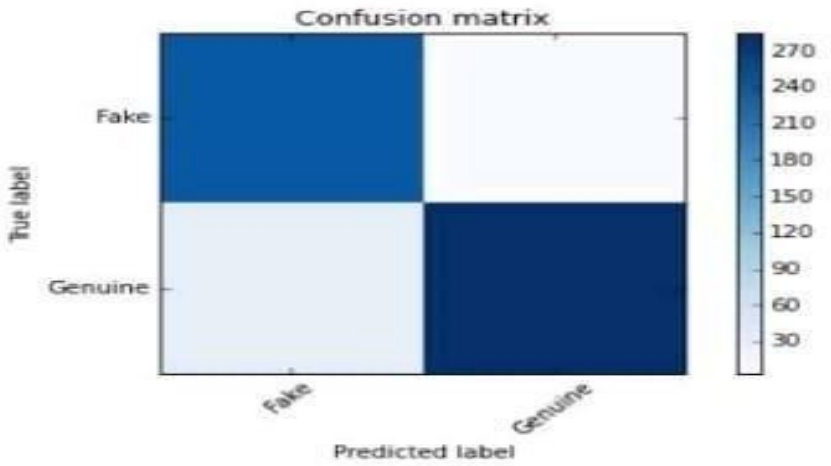

Figure 2: After doing the normalization of the confusion matrix.

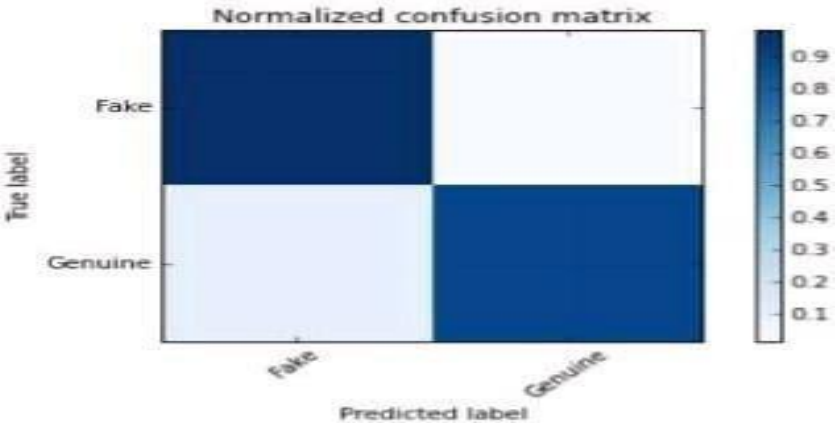

Figure 3: Normalized Confusion Matrix

Step 13: The classification of reports has been presented to define the fake and genuine profile precision index, recall, f1-score and support vector.

Step 14: Final outcome of the experiment is as following that describe the false positive and true positive values.

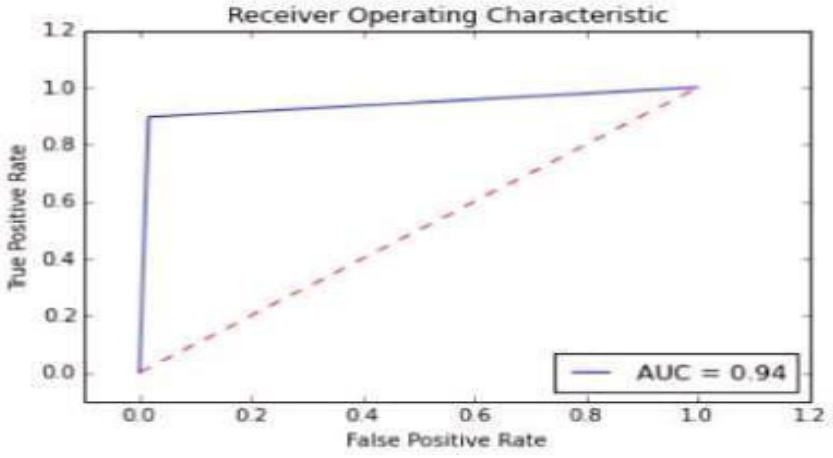

Figure 4: Receiver Operating Characteristic Curve

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Issue 4,June 2022

Experiments and Results

Procedure to detect Fake Profile using Support Vector Machine

Step 1: For the detection of fake profiles online social networks using SVM, several libraries need to be integrated including sys, csv, datetime, MATLAB and so on. These libraries are essential for reading the csv datasets and plotting the matrix.

Step 2: The second step contains, reading of datasets. In our case the name of the file is fusers. csv and users. csv to train and test the model. The genuine users are stored in user. csv file while fake or bogus users are in fusers.

Step 3: Define a function to get the information about the gender through name given in the dataset.

Step 4: For feature extraction, we have declared the function extract features.

Step 5: Further to this, we will draw the plot of learning curve which will provide information about certain features used to process the vectors.

Step 6: Next, there is confusion plot matrix associated with Fake & Genuine user profiles. Also, we will set the color value for plotting the same.

Step 7: For plotting the ROC or receiver operating characteristic, function has been defined.

Step 8: For training the dataset with support vector machine, function with the name of train has been declared along with the SVM classifier.

Step 9: Next step is for reading and extracting the datasets features.

Step 10: Different values will be shown after extracting the features, including count of statuses, follower counts, listed count, language code and so on for further processing.

Step 11: Splitting the datasets from the training and testing datasets.

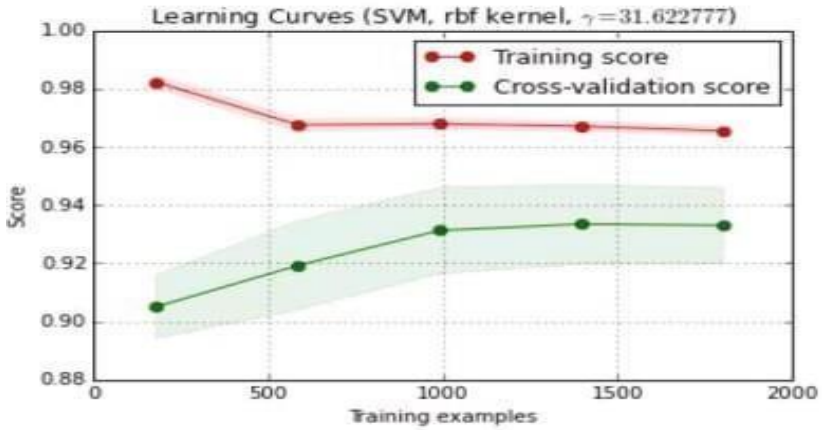

Figure 5: Training Examples

Step 12: The training data of learning curve will be displayed in red color and the cross-validation in green color. The score data has been mentioned along with the training.

Step 13: Predictive labeling of confusion matrix is being performed before normalization.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Issue 4,June 2022

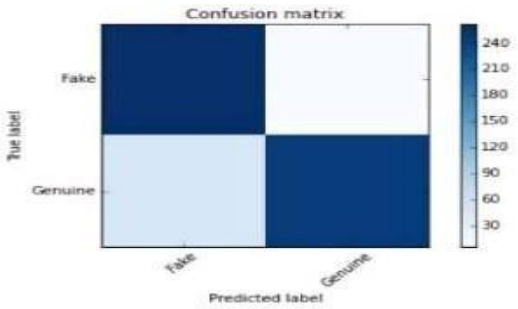

Figure 6: Confusion Matrix

Step 15: Lastly, plotting of ROC curve having True Positive and False Positive characteristics.

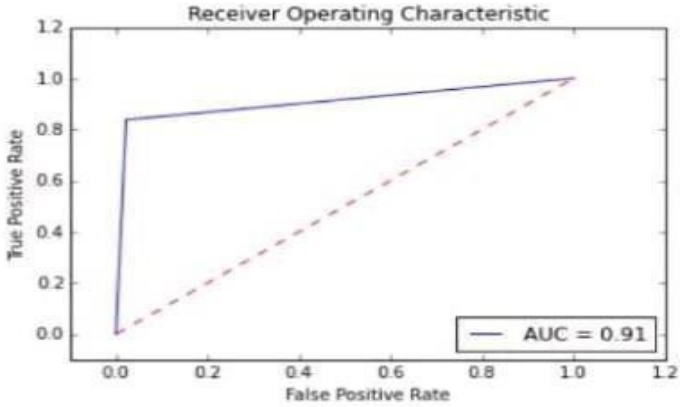

Figure 7: Receiver Operating Characteristic Curve

Conclusions and Future Scope of Work

Social Networks are booming in almost all the industries and they are becoming the main platform for companies to showcase their products or services to the end customers. Public Relation companies makes millions with social networks by publicizing the content related with different entities which can be political parties, any celebrity, institutions etc. Fake News with the help of fake profiles is also increasing at rapid pace and people use fake identities on social networks to publicize fake news, they are also related with the fake reviews, comments etc.

Social Network giants like Facebook, Twitter they continuously try to reduce the fake accounts by detecting them, but the problem is in actual keeps on rising with the rise of social network. We have used machine learning with python in order to detect the fake social profiles.

Three different algorithms i. e. Support Vector Machines (SVM), Neural Network (NN) are used and it is found that NN brings higher AUC than SVM. Researchers are continuously working on reducing if not eliminating this big problem on social networks and with continuous improvements in Artificial Intelligence features, researchers in future expect to minimize this issue.

References

[1] Breuer, A. , Eilat, R. , Weinsberg, U. (2020, April). Friend or Faux: Graph- Based Early Detection of Fake Accounts on Social Networks. In Proceedings of The Web Conference 2020 (pp. 1287- 1297).

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Issue 4,June 2022

[2] Balaanand, M. Karthikeyan, N. Karthik, S. Varatharajan, R. Manoharan, G. Sivaparthipan, C. B. (2019). An enhanced graph based semi-supervised learning algorithm to detect fake users on Twitter. The Journal of Supercomputing.
[3] Jiang, X. , Li, Q. , Ma, Z. , Dong, M. , Wu, J. , Guo, D. (2019). Quick Squad: A new single- machine graph computing framework for detecting fake accounts in large-scale social networks. Peer-to Peer Networking and Applications, 12(5), 1385-1402.
[4] Sahoo, S. R. , Gupta, B. B. (2020). Real-Time Detection of Fake Account in Twitter Using Machine-Learning Approach. in Computational Intelligence and Communication Technology (pp. 149- 159). Springer, Singapore.
[5] Pakaya, F. N, Ibrohim, M. O. , Budi, I. (2019, October). Malicious Account Detection on Twitter Based on Tweet Account Features using Machine Learning.
[6] Yuan, D. , Miao, Y. , Gong, N. Z. , Yang, Z. , Li, Q. , Song, D. ,... Liang, X. (2019, November). Detecting fake accounts in online social networks at the time of registrations. In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (pp. 1423-1438).
[7] El-Mawass, N. Honeine, P. Vercouter, SimilCatch: Enhanced social spammers detection on Twitter using Markov Random Fields. Information Processing Management, 57(6), 102317.
[8] ESTEE VAN DER WALT and JANELOFF Using Machine Learning to Detect Fake Identities: Bots vs Humans" Received December 5, 2017, acknowledged January 12, 2018, date of production January 23, 2018, date of current rendition Walk 9, 2018.
[9] Campos, G. F. , Tavares, G. M. , Igawa, R. A. , Guido, R. C.: Detection human, legitimate bot, and malicious bot in online social networks based on wavelets. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM).
[10] Cody Buntain, Jennifer Gol beck, "Automatically Identifying Fake News in Popular Twitter Threads", IEEE International Conference on Smart Cloud.