Detection of Fake Profiles on Online Matrimony Using Machine Learning and Deep Learning

Sneha B Perla[1] Tejal P Sarvade[2] Balram C Chaurasiya[3]
[1,2,3]Department of Information Technology in Engineering
[1,2,3]Padmabhushan Vasantdada Patil Pratishthan's College of Engineering, Mumbai, India

Abstract— Our paper focuses on the automation of the task in the field of matrimonial website. In our paper we created and demonstrated an idea to automate the task such as verifying profile, validating the documents provided by users and detection of fake profiles. This features are implemented with the help of Artificial Intelligence/Machine Learning algorithms combined together to avoid any intervention of humans in the task and provide a fully automated website. In this paper we represented our idea for detection of fake profiles on online matrimony sites by using machine learning and deep learning. We created a website of matrimony and on login page user has to upload a document and by OCR (optical character reorganization) with the help of this the data is extracted from image document and matches with data given by user. By using machine learning by anomaly detection the user is genuine or not we can find. Due to lack of labelled examples for in-genuine users, we solve the above problem as anomaly detection problem. In this thesis, we use autoencoder which is widely used algorithm for anomaly detection. We capture user's behavior, profile information and edit history to predict him/her as in-genuine or genuine profile. We then treat this problem as a reconstruction task using autoencoder which is trained on a set of genuine profiles features.

Keywords: Detection of Fake Profiles on Online Matrimony Sites, Deep Learning, Machine Learning and OCR

I. INTRODUCTION

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Because of new computing technologies, machine learning today is not like machine learning of the past. It was born from pattern recognition and the theory that computers can learn without being programmed to perform specific tasks; researchers interested in artificial intelligence wanted to see if computers could learn from data.

The iterative aspect of machine learning is important because as models are exposed to new data, they are able to independently adapt. After posting an exciting online matrimony profile on any reputed matrimonial site, posing as prospective bridegroom, fraudsters befriend women. They tactically use voice-changing apps to pose as the parents and guardians of the bridegroom when talking to the women they are trying to scam. Once they gain assurance, the fraudsters ask women to transfer money into their bank accounts expressing an emergency. They disappear without a trace as soon the money is transferred and this process continues for the next victim. There equal number of cases where bride is duped as well.

Many matrimonial websites have verified profiles that have been thoroughly background checked by and verification experts. Ideally you can progress conversation forward with a person having a verified symbol. Marriage is a lifetime decision. If you feel the other person is forcing you to take things forward quickly, be firm to take a back step.

II. RELATED WORK

The profile you liked on online matrimony sites could turn out to be your better half. Hence, it is imperative that you do a thorough profile check. Have a look at where they stay, their education, parents and work place and see if the story told online matches the background check. Visit their place of work. Do a thorough personal reference check? You should also look at their social media profiles. The functionalities provided by matrimony side are user register and login to profile, view profile update profile, capture image and upload photos, send request and accept request, chat with another user .Online matrimony is a hub for the users searching for their desired life partner. Unfortunately, it also attracts users with malicious intentions to come on portal and spam peer users. Apart from Online matrimony, the trend of online spamming is prominent on other web forums like dating sites. Conduct study of scams/frauds done on an online dating website.

They also proposed a taxonomy for different types of scammers present on an online dating website. Claims that bots are responsible for 51.8% traffic on online dating sites. They also conclude that these bots are very hard to differentiate from genuine users. As the functionality of online dating sites and online matrimony are quite similar, thus the existence of fake users is also a serious problem in matrimony domain. Thus there is a need to build a machine learning based system which can detect such fake 1 users on online matrimony. Different researches have contributed to detect fake users across different web domains. Hussain et al. [1] suggest a machine learning based framework to detect spam users in location based social networks (LBSN). The seriousness of this problem has also motivated researchers to identify spammer groups on social networking site like Facebook. Hsu et al. [2] tried to detect spammers who have created Facebook groups to spread misinformation. They have included the relationship between members and characteristic of their activities in the feature set. They then trained a support vector machine to detect spammer groups on the portal. [3] also propose a machine learning system which detects fake user accounts on Facebook. Based on user profile activities and interaction with other users, they have developed a feature set which when fed into a machine learning classifier produces 79% accuracy. Adikari et al. [4] have detected fake profiles on the LinkedIn dataset using machine learning algorithm. They claim that support vector machine with polynomial kernel outperforms other classifiers on the features captured from the LinkedIn dataset.

III. METHODOLOGY

On online matrimony the information provided by user may be fake or may not be fake. The extensive use of fake information has negative impact on online matrimony. The aim of this project is to find the fake information provided by the fake users using machine learning. One such fake profile which exhibits high inconsistency between attributes like religion (Hindu: Brahmin) and mother tongue (Urdu). Moreover, in this case education (High School), occupation (IT Engineer) and income (100,001 US dollars) do not comply with each other. A similar trend has for another fake user. Some profiles have been encounter.

The user can register himself and look for groom/bride and send a request to groom/bride .During the login page the user is verified by submitting his document we implemented ocr which it extracts the data from the image and then we match it with user given data if it matches then the user is not fake .During login user has to submit his Aadhar number should be provide by user and it should be valid if not valid then it will not allow user to go further .The user can send a request to another user if she or he accepts user request then they can have chats .User can't use abusive words and language if he or she tries to send abusive words to another user the website will not allow user to send such words. The user has to capture a image of him self for profile photo and three photos has to upload by using machine learning the face compare algorithm is used and this algorithm helps to detect the fake and genuine user with the help of this algorithm we match the face of user captured photo with his document submitted by user (pan card). If the face matches with 80% to 90% then the user is genuine and if doesn't matches then the user is fake user if user is fake then that profile is given alert warning and further it will be blocked by admin .

For an attribute like caste, religion, mother tongue, a Fake profile tries to send interest to most of the categories. On the other hand, a genuine profile tends to send interest only to those categories in which he is really interested in. This unique demarcation in pattern of sending interest becomes an important feature to judge if a user is authentic or not. So to capture user's behavior of sending interest in last 60 days of his activity we have designed behavior features in the following manner. It is evident that fake profile edit some crucial attributes like caste and mother tongue which ideally should not be changed in a user's lifetime. Moreover, the time difference between consecutive edits of a fake profile (for different categorical attributes) is very less whereas genuine profiles generally stick to one category for longer time. To capture this distinction in the edit pattern, we computed edit features in the following way . A unique characteristic of a fake profile is its inconsistency within profile attributes. There have been cases when a user claims to hold a Ph.D. degree at the age of 21 years. This accounts for profile inconsistency along the education and age attributes. Also, there exist inconsistency between the user's own category and the category to whom he/she is sending interests. We had an abundance of data of genuine profiles as compared to that of Fake ones. Thus the problem of class imbalance can arise while taking it as two class supervised learning problem.

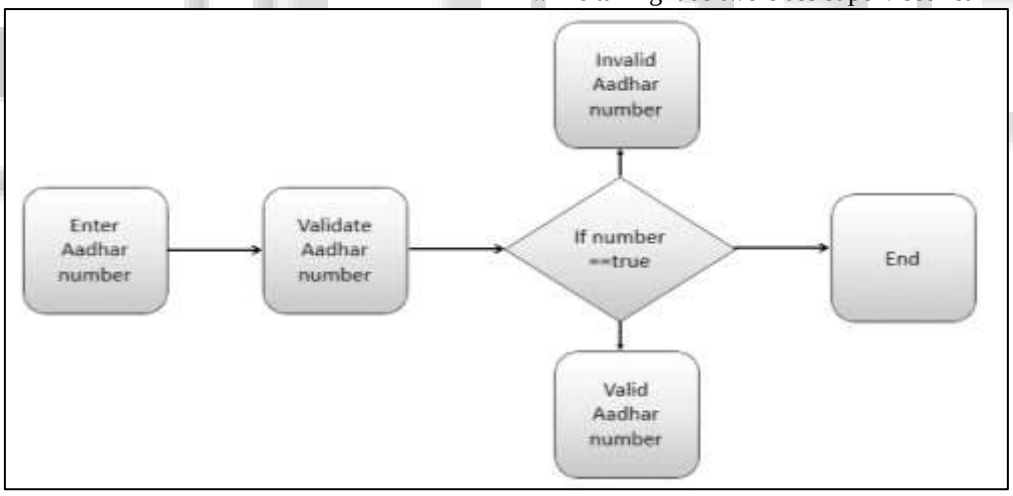

Fig. 1: Aadhar validation model

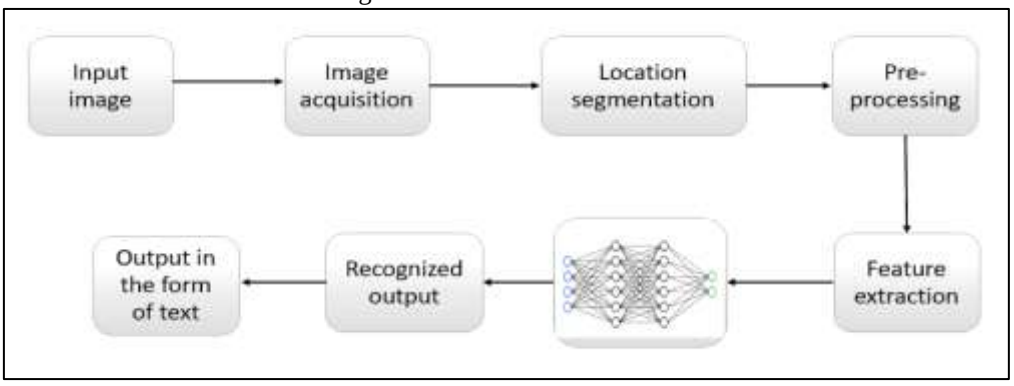

Fig. 2: OCR Model

IV. CONCLUSION AND FUTURE SCOPE

We first studied the distinction in behavior, profile and edit pattern between genuine and fake users. To detect fake profiles, we propose an anomaly detection method based on auto-encoders, for which the input feature vector is decided judicially. We describe the feature vector in the coming chapters. The proposed framework will save ample amount of time that were used to spend in manually detecting fake profiles.

Different researches have contributed to detect fake users across different web domains which suggest a machine learning based framework to detect spam users in location based social networks (LBSN). Big data analytics methods can lead to rapid detection of fraudulent profiles which may not be possible with the help of conventional techniques. The seriousness of this problem has also motivated researchers to identify spammer groups on matrimony sites like shaadi.com, jeevansaathi.com, etc.

REFERENCES

[1] Hussain A., Keshavamurthy B.N. (2019) AnalyzingâĂŸOnline Location-Based Social Networks for Malicious User Detection. In: Sa P., Bakshi S., Hatzilygeroudis I., Sahoo M. (eds) Recent Findings in Intelligent Computing Techniques. Advances in Intelligent Systems and Computing, vol 707. Springer, Singapore.
[2] Fu-Hau Hsu, Meng-Jia Yan, Kai-Wei Chang, Chih-Wen Ou, Hung-Min Sun. Itus: Behavior-based Spamming Group Detection on Facebook. In: Airiti Library 10.3966/199115992018082904006.
[3] Gupta, Aditi Kaushal, Rishabh. (2017). Towards detecting fake user accounts in facebook. 1-6. 10.1109/ISEASP.2017.7976996.
[4] Shalinda Adikari and Kaushik Dutta. Identifying fake profiles in linkedin. In 18th Pacific Asia Conference on Information Systems, PACIS 2014, Chengdu, China, June 24-28, 2014, page 278, 2014.

