International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

IDENTITY DECEPTION DETECTION ON SOCIAL MEDIA PLATFORM

Vindhya G B[1], Santosh J[2], Sumanth N R[3], Yashaswini N[4], Shreelakshmi M M[5]

[1]Assistant professor, Dept. of Computer Science Engineering, Sapthagiri College of Engineering, Bangalore, India
[2-5]Students, Dept. of Computer Science Engineering, Sapthagiri College of Engineering, Bangalore, India

---

Abstract - *Almost everyone in the present world use social media platforms for communication example twitter and Facebook. Social media platforms allow billions of individual users to share their thoughts, likes and dislikes in real time, that is without any censorship. Cyber threats are more difficult to detect in the social media platforms as there are many false identities and fake profiles are present. This project focuses on detecting those deceptive accounts by using machine learning techniques. Through which we can take further steps to minimize the rate of identity deception.*

1. INTRODUCTION

In the present world people are exposed to all sort of abuses on social media. The malicious intent of humans copying other humans identity constitutes a cyber threat that is one of the most difficult to deal with. As a result of cyber threats and SMP vulnerabilities are been increasing. By the word vulnerabilities we mean cyberbullying- those people who use social media and send abusive messages. Spammers- those people who intent to advertise their product or website by adding links along with the trending topics in the social media. In case of identity deception, a deceptive account which is created with malicious intent, which pose threat to other humans at large. A deceptive account with malicious intent could be for example be used to defame others. These deceptive accounts are generally created by humans or bots. Bots account require no human involvement for the action they perform. These deceptive bot accounts are know to target groups, as opposed to individuals.

- ★ To identify and collect all the attributes available like (date on which account was created, latitude and longitude of the tweet)
- ★ To experiment with these attributes by using machine learning models.
- ★ To enhance all the attributes to intelligently detect the deceptiveness of the users account.

This project also uses features derived from the field of psychology (emotions) like what they like and dislike, number of followers, whom they follow, what posts are they updating in their timeline on SMPs

Thus, to over all these vulnerabilities we develop a system called as "identity deception detection on social media platform".

2. LITERATURE SURVEY

In [1] In this paper they have considered the problem of detecting spammers on Twitter. They have first collected a large dataset of Twitter that includes more than 54 million users, 1.9 billion links, and almost 1.8 billion tweets. Using tweets related to three famous trending topics from 2009, Spammers post tweets containing typical words of a trending topic and URLs and lead the users to unrelated websites. They have used SVM (support vector machine) to classify spammers and non-spammers. And achieved a result by 70% accuracy on spammers and 96% on non-spammers.

In [2] The K-fold Cross Validation (KCV) technique is one of the most used approaches by practitioners for model selection and error estimation of classifiers. The KCV consists in splitting a dataset into k subsets; then, iteratively, some of them are used to learn the model and classify which model is more reliable, while the others are exploited to assess its performance.

In [3] In this paper, they suggest a list of high-level features and study their applicability in detection of cyberpedophiles. Pedophiles are people who create fake accounts and involve in child abuse. They have used binary text classification for classifying pedophiles and non-pedophiles by using behavior and chat conversation of a person as attributes.

In [4] In many Twitter applications, developers collected only a limited sample of tweets and a local portion of the Twitter network. Given such Twitter applications with limited data, they developed a collection of network-, linguistic-, and applicationoriented variables that could be used as possible features and identify specific features that distinguish well between humans and bots.

In [5] This paper focuses efficient detection of fake Twitter followers.

People who wants to have more followers, create many fake accounts and follow their own account by using those fake accounts and pretend to have more followers. It provides efficient techniques for detection of those fake

Twitter followers. They have used classification algorithms like random forest which gave more accuracy and with respect to cost naïve Bayes algorithm was found to be good.

In [6] This paper is based on deception detection on OSN (Online social Networks). They have used some of the attributes like gender-based classification and location-based classification, and they classify fake by using ML algorithms.

In [7] This paper discusses approaches and environments for carrying out analytics on Clouds for Big Data applications.

It revolves around four important areas of analytics and Big Data, namely (i) data management and supporting architectures; (ii) model development and scoring; (iii) visualization and user interaction; and (iv)business models. Through a detailed survey, we identify possible gaps in technology and provide recommendations for the research community on future directions on Cloud-supported Big Data computing and analytics solutions.

In [8] This paper is based on Recognizing predatory chat documents using semi-supervised anomaly detection.

They have utilized a new semi-supervised approach to mitigate this problem by adapting an anomaly detection technique called One-class Support Vector Machine which does not require non-predatory samples for training.

In [9] This paper is based on Cybercrime detection in online communications.

They have proposed set of unique features derived from Twitter; network, activity, user, and tweet content, based on these feature, we developed a supervised machine learning solution for detecting cyberbullying in the Twitter.

In [10] This paper is based on the, Inside the Black Box: How to Explain Individual Predictions of a Machine Learning Model. This paper contains several explanation methods which are described and compared on multiple datasets (text data, numerical), on classification and regression problems.

3. RELATED WORKS

The proposed work uses attributes found in person's Twitter account to detect human identity deception. Twitter is the only platform where no permission from the account holder is required to gather data. Comparatively, on Facebook, permission is required from each person before their data can be gathered, that is in the form pf friend request. Due to accessibility of its data, Twitter has been used.

A corpus was created by gathering data from Twitter, then cleaning it. The next step is involved in using the corpus as input to train machine learning models in detecting identity deception.

Two experiments are being defined:

Experiment 1: Data from the corpus is used to detect identity deception. This dataset was based on the original attributes as found in Twitter.

For example, twitter id, followers count, verified.

Experiment 2: The attributes used in Experiment 1 were then extended with new engineered features. The intention of the second experiment was to evaluate whether these features could possibly improve the accuracy of identity deception detection by humans on SMPs.

4. METHODOLOGY

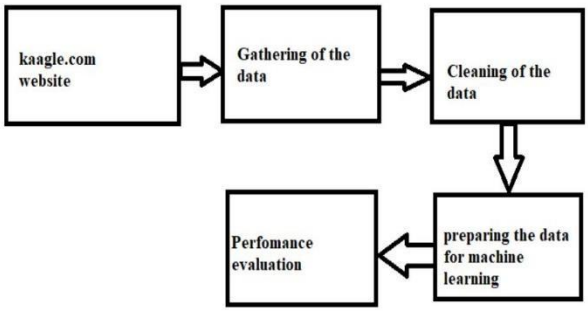

Figure: block diagram

Kaagle.com is a website which allows to find and publish datasets, explore and build model in and publish datasets, explore and build model in a webbased data-science environment. All the required data were collected from this website.

Gathering the data

Exactly 20,985 twitter account details were collected or gathered from the above website. All these twitter accounts which were created between 2015 to 2016, that is in the span of 1 year.

Cleaning the data

In cleaning of the data, Normalization technique has been used. Data normalization is a technique of organizing the data in the database. Normalization is a systematic approach of decomposing tables to eliminate data redundancy (duplication of accounts with respect to this project) and undesirable characteristics like insertion, update and deletion Anomalies. The Normalization technique finds the corelation between the different

accounts and if any accounts details are stored twice, it will be deleted.

Normalization technique also deletes the account details if any blank space in found with respect to attribute values, amongst the collected data.

Detecting identity deception

After the experiment 1 is done with respect to Gathering and Cleaning of the data. The final dataset will be passes to experiment 2.

In experiment 2, different machine algorithms will be used and will be trained with respect to- the engineered features (attributes). The algorithms which are used in our project are:

Knn algorithm

It is a supervised machine learning algorithm. In the random k value is chosen, and distance is calculated between the query and the current examples. Euclidian distance formula is used. The we sort the ordered collection, then we should pick the k entries. Based on that we can classify.

Random Forest

Random forest algorithm m works on the concept of large collection of decision trees. And it based on bagging technique. If we have sample A and B, where A is the first sample and B is nth sample. From these sample we create many random subsets with random values. Based on these sub decision trees class prediction is made to classify.

Ada-Boost Algorithm

It is a boosting algorithm, it is used to boost the performance of the machine learning algorithm, decision trees. Ada boost is more preferred in classification rather than regression process. It used to get high accuracy.

Naive Bayes algorithm

Naïve Bayes classifier is based on the concept of Bayes theorem, with the assumption of independence among the predictors. So generally, it is used in very large datasets.

Attributes

The table shows different attributes which are used of twitter which are termed are engineered features in this project. These attributes will be assigned unique values.

For example, consider the verified attribute. For this attribute value 1 0r 0 will be assigned. Among the collected data if the attribute values in 1 means the account is verified by the twitter and its account of some famous person.

twitter id
followers_count
friends_count
geo_enabled
default_profile
default_profile_image
time_zone
protected
verified
notifications
follow_request sent
status
listed_count
contributors_enabled
followings
is_translator
profile background image

Table: twitter attributes

So, it is authentic. If its 0 it comes under the list of normal people.

Likewise, each attribute attributes contributes in finding the deceptiveness in the account.

Attribute	Value
NAME	Marco Duran
SCREENNAME	tinyfrog537
CREATED	3/6/2016
PROFILE_IMAGE	https://randomuser.me/api/portraits/women/99.jpg
LOCATION	Las Palmas De Gran Canaria
LANGUAGE	En
FRIENDS_COUNT	14
FOLLOWERS_COUNT	924
STATUS_COUNT	1 670
LISTED_COUNT	409
TIMEZONE	America/New York
UTC_OFFSET	-18 000
LATITUDE	85.03769
LONGITUDE	-136.27587

Table: example of a deceptive account

The above table is the best example of a fake account. This account is recognized as fake due to a number of

reasons, the name and screenname are not related, the image represents a different gender than suggested by the name, the latitude and longitude coincide somewhere over the Arctic ocean, and New York's UTC offset is actually -4 and not -18 as suggested.

The existing system there is no automated system to detect the deception. Accuracy is less than 50% in detection of deception accounts in online social media. In our project we calculate:

MAE (Mean absolute error)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|$$

Where, n = total number of points y = actual output value y^= predicted output value y^-y=the absolute value of residual

MSE (Mean square error)

$$MSE = \frac{1}{n}\Sigma\left(y - \hat{y}\right)^2$$

Where,

(y-y^)2 = the square of the difference between the predicted and actual value.

R squared value

$$R^2 = \frac{SSR}{SST}$$

Where,

SSR = sum square regression error

SST = sum square total error

RMSE (Root mean square error)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Where, n = total number of points.

So, we calculate all these values for each of the machine learning algorithm and compare the result.

	MAE	MSE	R2	RMSE	Accu
RF	0.0002	0.0002	0.998	0.0154	99%
AdaB	0.0	0.0	1.00	0.0	100%
NB	0.272	0.2726	-0.175	0.5221	72%
Knn	0.202	0.202	0.128	0.449	79%

Figure: Result

5. CONCLUSIONS

Cyber security can benefit from this project which involves in development of identity deception detection by means of machine learning models. By using different attributes machine learning algorithms are trained. And the result are been compared. In our project Ada boost algorithm showed higher accuracy with 100% with high r square value, then next best algorithm was random forest algorithm with 99% accuracy, then followed by knn, naïve Bayes showing less accuracy.

Future research work can be to identify more features which are valuable towards the detection of identity deception detection on SMPs including the refinement of the model.

REFERENCES

1. [Al-Garadi MA, Varathan KD, Ravana SD. Cybercrime detection in online communications: the experimental case of cyberbullying detection in the Twitter network. Comput Hum Behav 2016;63:433–43].
2. [Alowibdi JS , Buy UA , Philip SY , Ghani S, Mokbel M. Deception detection in Twitter. Soc Netw Anal Min 2015;5:1–16]
3. [Anguita D, Ghelardoni L , Ghio A , Oneto L , Ridella S . The 'K'in K-fold cross validation. Proceedings of the European symposium on artificial neural networks, computational intelligence and machine learning; 2012. p. 441–6]. [Beillevaire, M. 2017. *Inside the Black Box: How to Explain Individual Predictions of a Machine*

4. *Learning Model.* Computer Science and Engineering Masters, KTH Royal Institute of Technology] [Assunção MD, Calheiros RN , Bianchi S , Netto MA , Buyya R . Big data computing and clouds: trends and future directions. J Parallel Distrib Comput 2015;79:3–15 .]
5. [Benevenuto F , Magno G , Rodrigues T , Almeida V . Detecting spammers on twitter. Proceedings of the collaboration, electronic messaging, anti-abuse and spam conference (CEAS); 2010. p. 12] . [Cresci S , Di Pietro R , Petrocchi M , Spognardi A, Tesconi M . Fame for sale: efficient detection of fake
6. Twitter followers. Decis Supp Syst 2015;80:56–71] [Caspi A, Gorsky P . Online deception: prevalence, motivation, and emotion. Cyber Psychol Behav 2006;9:54–9 .]