FAKE ACCOUNT IDENTIFICATION AND ACCOUNT CATEGORY PREDICTION IN TWITTER USING MACHINE LEARNING

PROJECT REPORT

submitted by

AFSAL A K (TVE16CS004)

AMRITHA S B (TVE16CS013)

ASHARUDHEEN T (TVE16CS022)

VISHNU U V (TVE16CS062)

to

the APJ Abdul Kalam Technological University

in partial fulfillment of the requirements for the award of the Degree

of

Bachelor of Technology

in

Computer Science and Engineering

Department of Computer Science and Engineering

College of Engineering, Trivandrum

Kerala

June 17, 2020

DECLARATION

I undersigned hereby declare that the project report **FAKE ACCOUNT IDENTIFICATION AND ACCOUNT CATEGORY PREDICTION IN TWITTER USING MACHINE LEARNING**, submitted for partial fulfillment of the requirements for the award of degree of Master of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of **Prof. Sreelal S**. This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Place: Trivandrum

Date: June 17, 2020

Afsal A K
Amritha S B
Asharudheen T
Vishnu U V

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
COLLEGE OF ENGINEERING, TRIVANDRUM

CERTIFICATE

This is to certify that the report entitled **"FAKE ACCOUNT IDENTIFICATION AND ACCOUNT CATEGORY PREDICTION IN TWITTER USING MACHINE LEARNING",** submitted by **Afsal A K**, **Amritha S B**, **Asharudheen T**, **Vishnu U T** to the **APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY** in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide record of the project presented by them under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Prof. Sreelal S	**Dr. Saritha R**	**Dr. Salim A**
Assistant Professor	Assistant Professor	Professor
Department of CSE	Department of CSE	Department of CSE
(Guide)	(Coordinator)	(Head of Department)

ACKNOWLEDGEMENT

First and foremost, we wish to place on records our ardent gratitude to our project guide Prof.Sreelal S,Assistant Professor,Dept. of Computer Science and Engineering.His tutelage and guidance was the leading factor in translating our efforts to fruition.We are extremely happy to mention a great word of gratitude to Dr.Salim A, Head of the department for providing us with all facilities for project. Finally yet important, we would like to express gratitude to project coordinator Dr. Saritha R for her valuable assistance provided during the course of project. We would also extend our gratitude to all the staff members in the department. We also thank all our friends and well-wishers who greatly helped us in our endeavor.

Afsal A K
Amritha S B
Asharudheen T
Vishnu U V

ABSTRACT

In an era where social medias are preferred over other traditional medias for news and information, the spread of a simple fake news can make it true. Many parties use this method to spread fake news through fake accounts as they have no source and can be anonymously operated as well. Also cyber bully are increasing day to day by the use of fake accounts as well. So here we propose a way to detect these fake accounts In social medias like Twitter effectively so that people can ignore their posts and the info shared by them as well. We implement an application which detects whether a given account is fake or not by the use of feature extraction and machine learning to categorise them as fake or not and to predict the category of the account as well.

Contents

List of Figures

List of Tables

ABBREVIATIONS

(List in the alphabetical order)

CART Classificatiion And Regression Tree

LDA Linear Discriminant Analysis

LR Logistic Regression

ML Machine Learning

NB Naive Bayes

SVM Support Vector Machine

NOTATION

(List in the alphabetical order)

A The area of the needle point

N The number of angels per needle point

a The number of angels per unit area

Chapter 1

Introduction

In this world of internet Online social network has become the indivisible part of our life. Some of the examples of these networks are Facebook, Twitter, WhatsApp, Instagram etc. One of the major problem faced by these platforms are cyber bullying and many types of false propoganda. The major cause of these problems are mainly fake account users. After seeing such fake news and atricles some may believe these and also they may retweet and this countinues. It will be better if we could find which one is true and which is fake and prevent the spread of these news. This will reduce the number of cyber crimes.

When new features are added to social networks, the number of fake account also increases. So its not easy to identify whether an account is fake or not. And also there is no such standardised way to , identify if its genuine. Many researches have been going through this topic. Detection of fake account become more tough when privacy to the network is more. In case of Facebook, it wont allow a third party to extract every feature from it. So without extracting much features fake account detection may not be that much easy.

One such major network is twitter. Many cases are reporting world-wide about the misuse of these online social networks. The purpose behind the creation and usage of such fake accounts in twitter varies. Some of them are in the names of celebrities and politicians for increasing the followers and for spreading fake news. similarly some increase the followers count to promote some product or someone. Another use is to fake reviews (paid reviews) and sponsored polls. So the people can't easily find the genuinity among the tweets and the twitter account.

So the main aim of this project is to build a web application for identifying weather the suspected twitter account is fake or not and to find the category of post the account is posting ,such as political,religious, advertisements etc. This web application accepts the user id of the twitter account as input and returns the account is fake or not and added to that predict the category of the account if it is fake.

The web application was created using Django web frame-work .Anaconda is used as the environment for windows. Twitter API, tweepy is used to get access to many twitter features to read the user profiles. Using this the user id of the suspected account is input in the web application. We used machine learning methods to classify whether it is fake or not. For machine learning features extracted from the account is used. Another machine learning method is used to categorize the fake account.

Support Vector Machine (SVM), Naive Bayes (NB) and Machine Learning (ML) Classificatiion And Regression Tree (CART) Linear Discriminant Analysis (LDA) Logistic Regression (LR) are abbreviations whereas a, N and A are part of the nomenclature

Chapter 2

Existing Methods

There have been several detection strategies developed to identify the problem in social networks. These problems are mainly due to fake accounts and fake news. One such platform is twitter. Among the existing methods use only a few features to identify an account is fake or not. So when the number of features are less, the accuracy also decreases and the efficiency falls.

2.1 Detection using Random Forest algorithm

The most common method for fake account detection is done through Random forest algorithm. It uses a number of features for classification, since feature reduction is not needed in this algorithm. There for no feature scaling required . The advantage of using this classification algorithm are, as the algorithm is based on the bagging algorithm, it reduces overfitting ,reduces variance and also less impacted by noise.
Even though so many advantages are there for Random forest algorithm, it has few disadvantages . The disadvantages are high complexity , as it creates several trees.So it require more computational power as well as resources .Another disadvantage is Long training period, as number of trees are high , time required to generate them will also increase. Also there is inefficiency to handle the categorical variables which has different number of levels.

Random forest algorithm accuracy 95 percent accuracy. But this accuracy and efficiency is reduced when new data come in a large amount.Not only random forest algorithm , but all existing method for identification of fake account may not be sufficient in this era.So according to current scenario of social networking, definitely its efficiency reduces and takes large time to predict whether an account is fake or not.

In this era of social networking, a huge number of fake accounts are created and used.Also a large number of new features are added by each social networking platform.Almost everything in social medias are continuously changing day by day.New datasets are produced in every minute.So the dataset is increasing drastically. So usage of Random forest algorithm is inefficient to handle such huge data. Another method is to be used for fake account detection. Another such method is discussed in chapter 4.

Chapter 3

Objective

This project mainly targets in making a difference in 2 areas. One of them is False propaganda and False news spreading using social media which strongly influences in Elections, people's mindsets as well. The next area in which this project targets is Cyber bullying and blackmailing in social medias. As we know these days the no. of fake accounts keep increasing and these causes many problems which will be discussed in further below.

3.1 FALSE PROPAGANDA

As we all know, many parties relies on social medias to popularize their propaganda and their good sides through social media, because of the fact that most of the youth and adults relies on reading news through social medias. So most of them won't care where the news came from or whether its a genuine one or not. So they would believe any news that an account provides. So here if we could verify the real source of that news is a fake account and using the category of the account we get from the application , we can verify if its a believable news or not. And since these propaganda plays a crucial role in elections and other methods. This can make an impact in selecting a government even.

3.2 CYBER BULLYING

Well one of the most common cases in these times for a cyber police is related to cyber bullying. Some random users who uses fake accounts try to get data of normal users and blackmails them with this data, attacking some random celebrity in their comment sections with this fake accounts , posting hate comments and posts about some one in some other groups and all. These all became most common now with the growth of social medias, and because of the fact that they uses fake accounts to do this type of dirty work makes them believe they are invisible.But as the proverb says , Truth always comes out , so here if we could use some sort of fake account identification method so that the user who checks for the genuinity of the account can report the fake accounts to the authority, we could prevent this types of wrong use of social medias as well. This is what exactly our project provides as well.

Concluding both of these main objectives , we could add much more points such as stopping the depressions people suffers because of these fake profiles and their works and help creating a better society which uses social medias in the better way where it works as some thing which connects the entire world and gives a complete freedom of expression for anyone in a way that it hurts no one as well.

Chapter 4

Design and Implementation

This chapter will be discussing the features used for design and the method for implementation in detail.For the web application designing and implementation, first we have to get some of the twitter user profile information. We gathered 60 such accounts, and using twitter API several features (total number of tweets, likes, followers , bio etc.) are extracted from each account.These extracted features are used to test and train the data. Some of the classification algorithms are used for thee training and testing of the data. And the algorithm with most accuracy will be selected for implementation.

4.1 ARCHITECTURE

In this approach we are focusing on an architectural methodology for implementation as shown in the figure . The system consists of several components, depicted as the few stages in the implementation and the concept here is to predict the suspected twitter account is Fake accounts or not and if the account is fake, the system also try to predict the which category of fake account that is like political, celebrities etc.For both prediction (fake or not and category prediction), different algorithms are used.Because fake or not prediction is based on some extracted feature values and category prediction is basically text classification.

Figure 4.1 shows the Components of the system

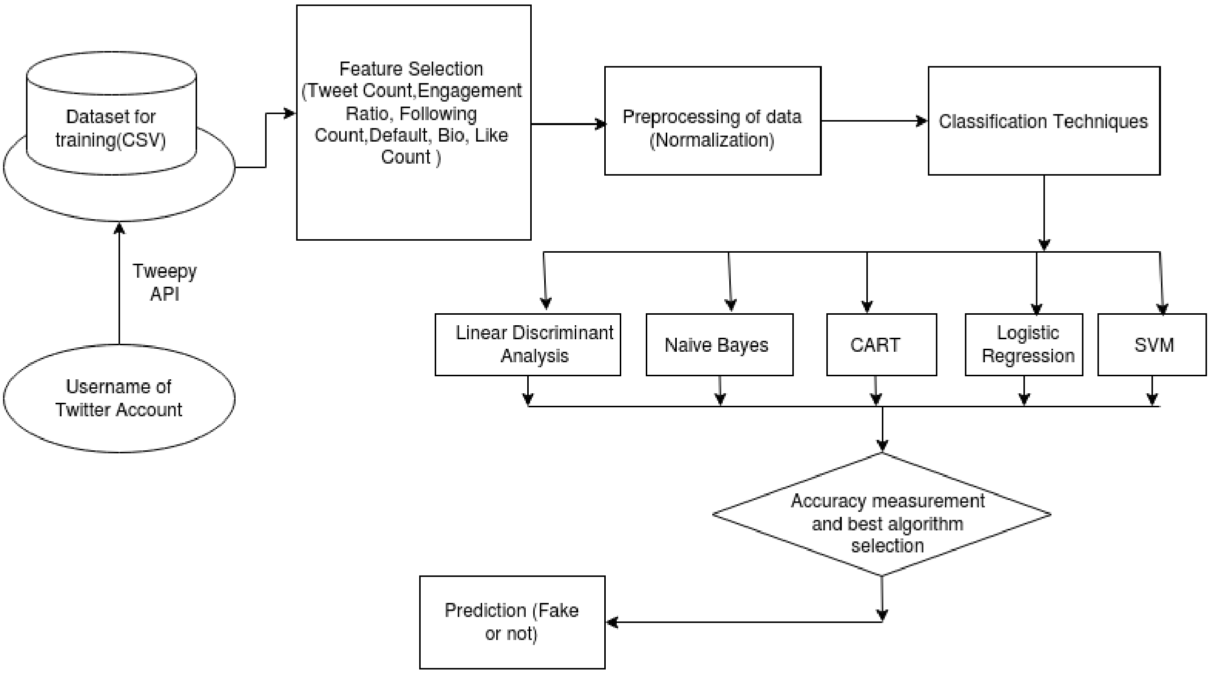

Dataset for training(CSV)
Feature Selection (Tweet Count,Engagement Ratio, Following Count,Default, Bio, Like Count)
Preprocessing of data (Normalization)
Classification Techniques
Tweepy API
Username of Twitter Account
Linear Discriminant Analysis
Naive Bayes
CART
Logistic Regression
SVM
Accuracy measurement and best algorithm selection
Prediction (Fake or not)

Figure 4.1: Architecture

(i) Username of twitter account

Every twitter account has its user id. This user id is username and this will be as the input to the system. When a suspected twitter account is found ,we can input its username in this application. Which will undergo the following processes and output us with the result as genuine or fake account (along with the category of the fake account).

(ii)Data set

We found 60 twitter account (both fake and genuine) manually for training and testing.From these twitter accounts several features are extracted . And this data set we used.

(iii)Feature selection

Huge number features are available in twitter. Among that features, a few features that are most suitable for fake account detection is used.Theses features are extracted directly from the twitter.The features include

- Total number of tweets

- Engagemnt ratio, Which is likes of a tweet/ Total number of tweets.
- Following count
- Length of user bio
- Count of likes for other users tweets
- Verified or not
- Default profile or not

(iv)Pre-processing of data

Before training the features that are extracted from twitter accounts (numerical values) need to be normalised. Normalization is important because the extracted features are not on the same range. Normalization ensures that there is no redundancy . So when normalization is done, the dataset will be in the same range and the processing of data ensure correct result.The normalised data is used in all other steps.

(v)Classification Techniques

From the preprocessed data we have to identify whether an account is fake or not. We used binary classification to identify whether an account is fake or not. The data is processed through five algorithms.They are discussed below.

- Logistic Regression(LR)

  It is a statistical model, that in its basic form uses a logistic function to model a binary dependent variable.Here LR can be used because the target is categorical, that is either fake or not (the data belong to any one class).

- Linear Discriminant Analysis(LDA)

  LDA is a dimensionality reduction technique. It minimizes the variance and maximizes the variance and maximizes the distance between the means of the two classes.

- Classification And Regression Tree (CART)

  CART output a decision tree where each fork is a split in a predictor variable and each end node contains a prediction outcome variable. In this scenario the end node is either Fake account or genuine account.

- Naive Bayes classifier (NB)
  This classification algorithm is based on Bayers theorem(finds the probability of the event). In this scenario prediction of event(Twitter account is fake or not) is done based on the features extracted from the corresponding account.

- support vector machine (SVM)
  SVM is also two-group classification algorithm. This algorithm uses kernel trick , that is, the algorithm transforms the data (extracted features from twitter account) and the based on that if finds the optimal boundary (the boundary for each extracted feature) between the outputs (account as fake or not)

(vi)Best algorithm selection

The next process is to select the best algorithm for classification.This is one of the crucial steps among the architecture. The data set training and testing is done with all the five algorithms. And the algorithm with highest accuracy can be selected for the implementation of the web application. Accuracy depend on the selected features for classification as well as the application.

(vii)Prediction as fake or not

Final step is to predict whether the suspected account is fake or not.By the testing through the best algorithm selected, we can identify the fake account and genuine account. If the account is fake then its category is also predicted.

For the category prediction,SVM is used. There are 20 category used , that is the fake account belong to any one category among such 20 categories. They are Atheism, Graphics,O.S, Hardware, windows, for sale, autos, motorcycles, sports, hockey, cryptography, electronics, medical, space, religion(Christian), politics(guns), middle east politics , other politics , other religions. Here more than 2500 data set are taken from internet for training and testing . 70 percent of the data set is used for training and 30 percent for testing.

4.2 IMPLEMENTATION

We created our web application, Fake or not using Django. Django is a widely used Python web application framework .Django frame work is simple and has many useful features for development.There are so many frequently used web application like Instagram,Spotify,NASA etc. are developed using Django. We choose Django to implement the web application because, as we discussed in previous section architecture, classification algorithms (ML Algorithms) are used. ML is highly supported by Python, and Python is supported by Django. we used ML libraries such as sklearn,pandas etc. We used Anaconda as the environment for windows. Anaconda supports multiple versions of Python and associated packages.

As we have described in architecture, the first step is taking the suspicious account username .That can be easily taken from twitter. And for further steps, the accounts features are to be retrieved. These features can be retrieved using Twitter API called tweepy (A Python library). This API is simply a set of URLs that take the parameters. Our 6 features can be retrieved using this. We chosen twitter to detect fake account because its access to data , which Facebook does not allow. Facebook does not allow third party use its users information.

The features used are six. First is the tweet count that is the total number of tweets the an account has tweeted and that can be directly taken using tweepy.Then Engagement ratio, which is total number of likes the user has got for the tweet in last week or probably the last tweet dived by the total number of followers. For a genuine user this number will be a large number but in case of fake account it will be very less. For example, if a person has 1000 followers he will get at leat 100 likes for a tweet and for fake user, the likes may be very few probably within 10.Third one is friend count , that is if a person follow someone and if that someone follow the person back, the person can be counted as friend. Usually a genuine account have follower count as well as friend count are similar and for fake account there may be huge difference in friend count and followers count.Next one is bio, its basically a self introduction about the user by user itself.For a genuine account it will be short, that is their bio length will be less and for fake account, they usually write a paragraph that are their bio length will be high.Next is like count , that is how much like the user has done for others tweets. Generally for a genuine account it may be a high number.Next is verified, some account are already verified as genuine that may be of celebrities, politicians , or some famous persons. If an account is verified it is indicated by a blue tick along with their user id. And verified account is already genuine and the feature we assumed may be exception for such accounts .So no need to check it again. Next is default. When a new twitter account is opened it will have a default theme. Generally a normal user changes their theme and then the default value will be 0.Otherwise default value will be 1.So these features are the data set.

The features are extracted from 60 twitter accounts (30 genuine and 30 fake). The next step is the normalization . As there is huge difference , normalization is done. After normalization is the training of the dataset (features) using different algorithms such as LR, LDA, CART, NB and SVM.For training we used 70 percent data. And remaiming 30 percent for teesting. Training with these algorithms will gives different result and the most accurate training algorithm is selected..After the training the data we can test. The suspicious account user id can be given as input and after testing it will result either genuine account or fake.

If the account is fake then that account's category also predicted.We used more than 2500 dataset which from internet for training and testing.For that prediction some features are used. The features are Count vectorizer is to count every word .Tfid-id Transformer is next feature it count similar words in similar category of data and it transform data from delimiters comma and other special characters .Then pipeline is implemented to decrease the time taken for training and prediction of data category. After training is testing. And the input is the users bio for predicting category of fake account.And this outputs the category.

Chapter 5

Results and Discussion

As discussed earlier we used 42 (both genuine and fake) twitter account for training . And the profile features extracted are listed below. They are Tweet count , engagement ratio , following count , length(bio) , total likes, verified or not(1 for true) , default profile or not(0 for true).

36911	0.161018	135	106	27936	0	1	fake
88975	0.040349	303	80	2526	0	0	fake
612	0.661376	283	65	867	0	0	fake
16488	1.635769	200	158	19346	0	0	fake
91664	0	709	115	22595	0	0	fake
121	166.6667	358	31	60	0	1	fake
2082	2.966843	58	112	193	1	1	genuine
50482	2.98765	47	48	6	1	1	genuine
5278	25.62978	947	46	106	1	1	genuine
369	43.9	939	13	359	0	0	genuine
30	62.5	371	0	3	0	0	genuine

Table 5.1: Dataset before Normalization

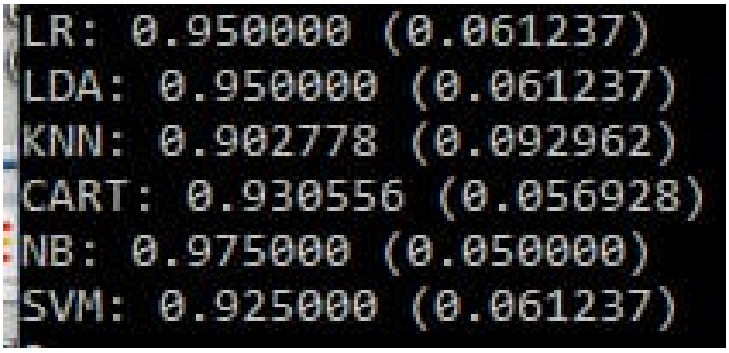
We can see that each value varies in a huge range. So they are normalized to make each feature value into same range.Thus the data changes as listed below.

0.075589	0	0.025681	0.607595	0.020956	0	0	fake
0.119285	0.000725	0.00642	0.670886	0.113475	0	1	fake
0.88975	0.000182	0.01441	0.506329	0.010261	0	0	fake
0.001978	0.002976	0.013459	0.411392	0.003522	0	0	fake
0.001649	0.007361	0.009512	1	0.078583	0	0	fake
0.296229	0	0.033719	0.727848	0.09178	0	1	fake
0.000391	0.75	0.017026	0.196203	0.000244	0	1	genuine
0.006728	0.013351	0.002758	0.708861	0.000784	1	1	genuine
0.163142	0.013444	0.002235	0.303797	2.44E-05	1	1	genuine
0.017057	0.115334	0.045037	0.291139	0.000431	1	1	genuine
0.001192	0.19755	0.044657	0.082278	0.001458	0	0	genuine

Table 5.2: Dataset after Normalization

5.1 ALGORITHM COMPARISON

The accuracy of each classification (prediction of twitter account as fake or not) algorithm is listed below.

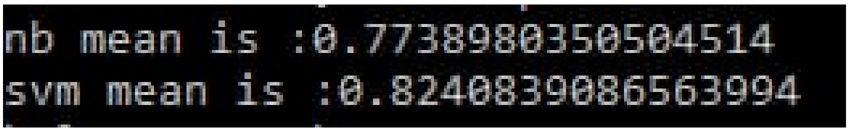


```
LR: 0.950000 (0.061237)
LDA: 0.950000 (0.061237)
KNN: 0.902778 (0.092962)
CART: 0.930556 (0.056928)
NB: 0.975000 (0.050000)
SVM: 0.925000 (0.061237)
```

Figure 5.1: Prediction(fake or not) algorithm accuracy

From this we can understand NB classifier has the highest accuracy that is 0.975000 (97.5%).So this is selected for the web application to detect whether a twitter account is fake or not.

Similarly for understanding the category (text classification), SVM classifier has 82% accuracy and NB classifier has 78% accuracy. So SVM is used for categorization.

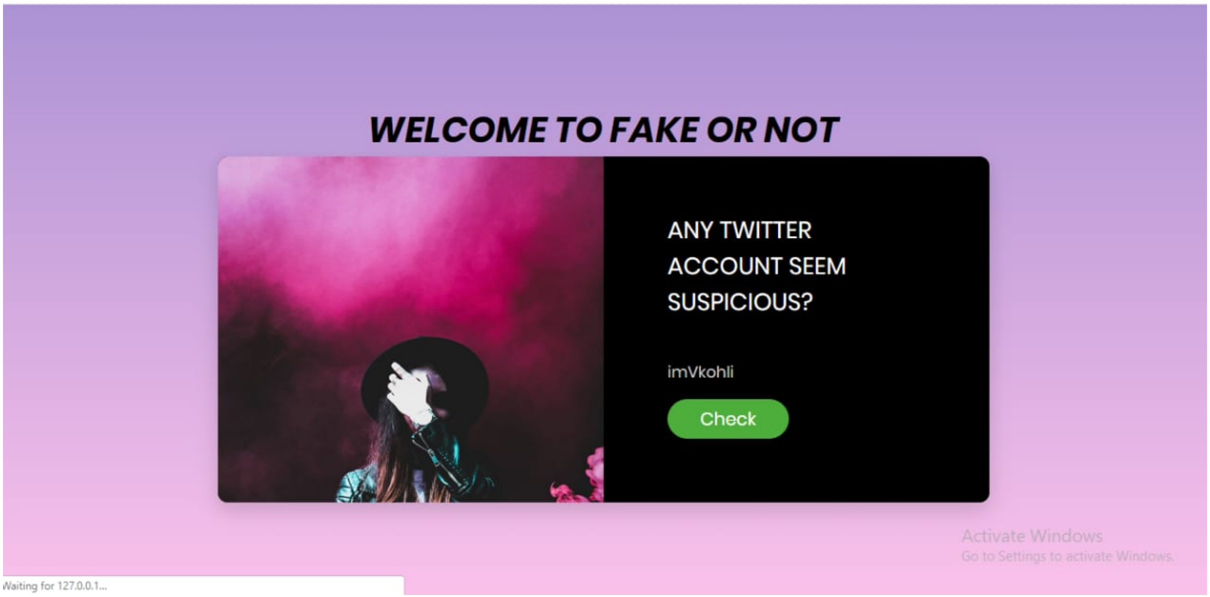


```
nb mean is :0.7738980350504514
svm mean is :0.8240839086563994
```

Figure 5.2: Categorization algorithm accuracy

5.2 THE WEB APPLICATION

The figure shows the how the application look and how it works. In figure 5.3 input is the suspected account user name as "imVkohli". After the input is given to the application it will train the input(username) using the implemeted algorithm.The result in figure 5.4 indicates the inputed account is genuine.

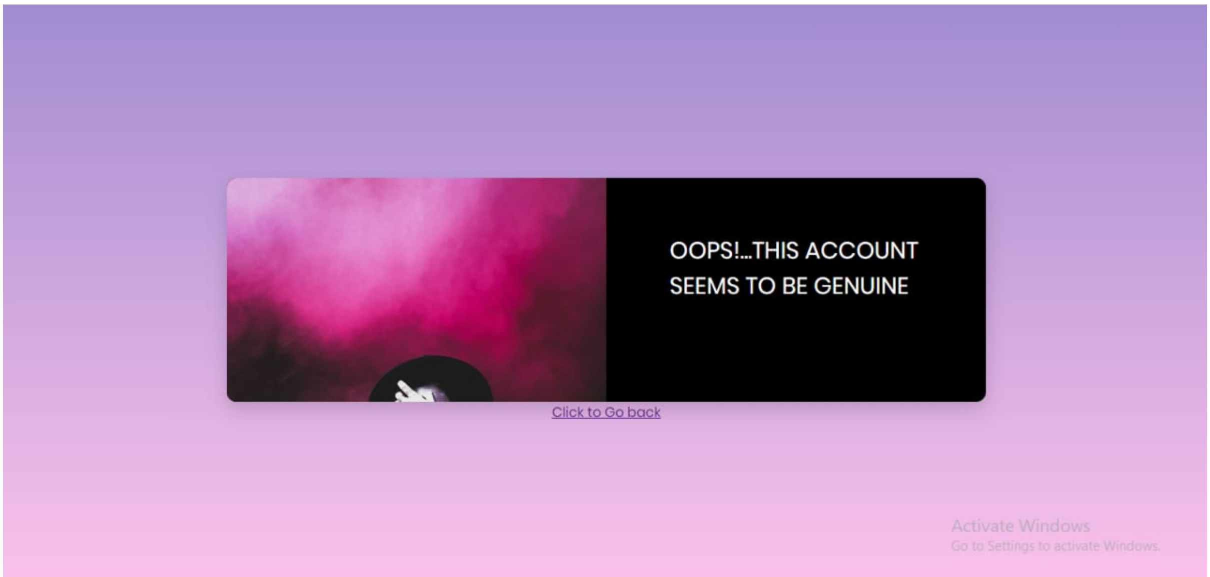


WELCOME TO FAKE OR NOT

ANY TWITTER ACCOUNT SEEM SUSPICIOUS?

imVkohli

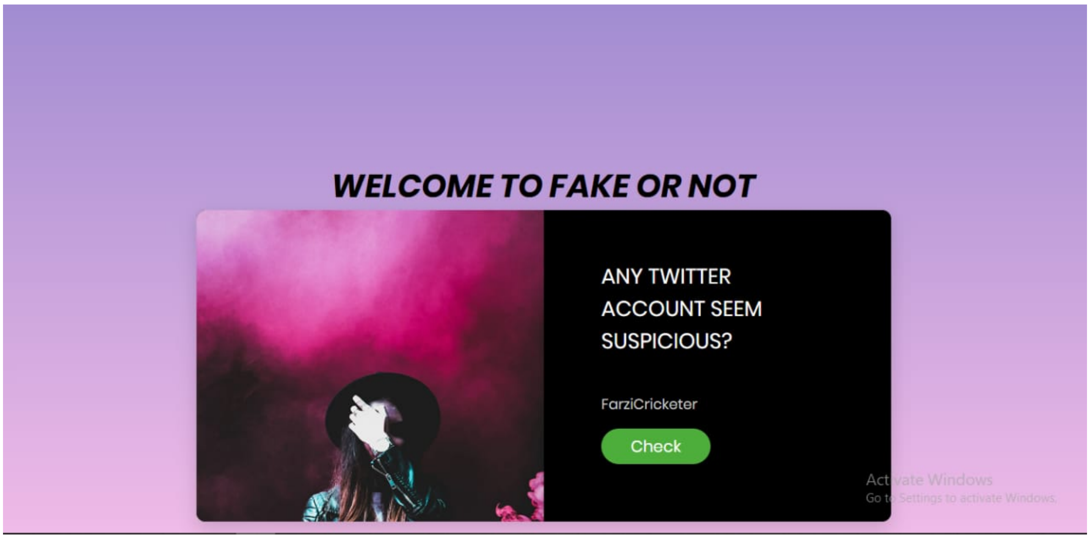
Figure 5.3: First input to the web application



OOPS!...THIS ACCOUNT SEEMS TO BE GENUINE

Figure 5.4: Result of fig 5.3

The figure 5.4 is the next input to the application as "FarziCricketer"(username of suspicious twitter account). From figure 5.5 its found that the suspicious account is fake. The category in which the account belongs to motorcycles. The category is predicted by categorization algorithm(SVM).

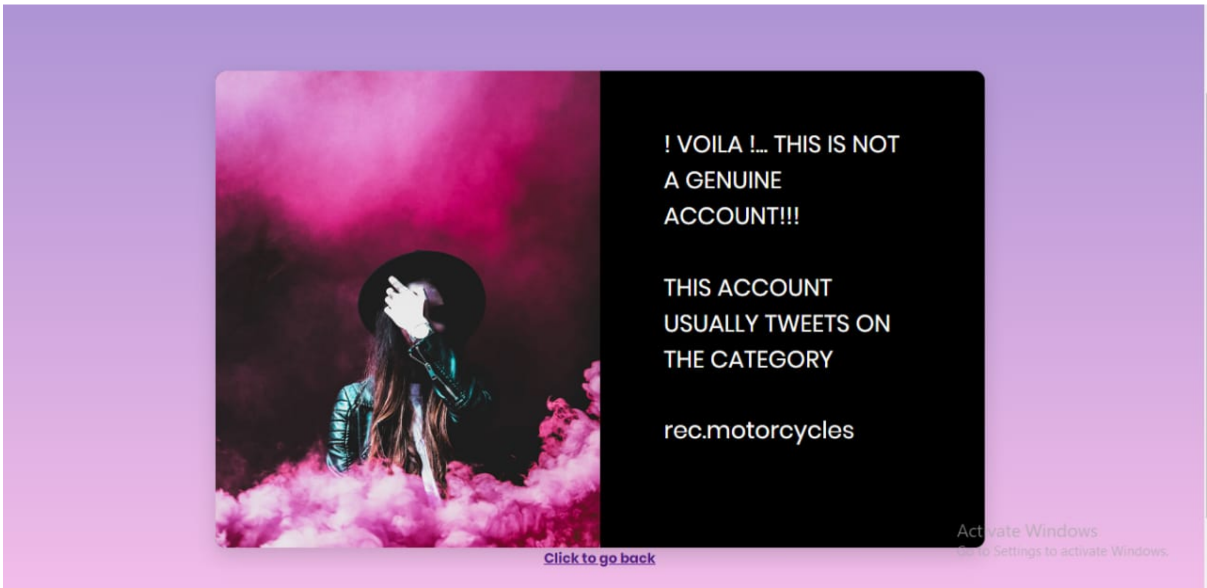


WELCOME TO FAKE OR NOT

ANY TWITTER ACCOUNT SEEM SUSPICIOUS?

FarziCricketer

Figure 5.5: Second input (username)



! VOILA !... THIS IS NOT A GENUINE ACCOUNT!!!

THIS ACCOUNT USUALLY TWEETS ON THE CATEGORY

rec.motorcycles

Figure 5.6: Result of input in fig 5.5

Chapter 6

Conclusion and Future Scope

Using the above proposed application we can reduce the increasing cuber bullying and fake news spreading effectively contributing to a better and safe experience of using social medias and cultivating its better sides the most as well, also preventing any one from taking steps into the dark sides of social media as well. By reporting these fake accounts to the administrators of the social medias, we could help them in reducing people using these fake accounts as well.

In the future, if it became possible we could link this application with the twitter platform so that these fake accounts are plucked out in their birth stage and making these sites secure as well.Not only in twitter, if we could make it acceptable by other social media authorities like facebook, instagram we could implement this in their applications as well and can be used in the whole social media world as well. Also if this concept is taken forward, we could really make miracles in elections and other politics as well, because people can always count on the information they get as they are genuine and something to share with others as well. This could be really helpful in democratic countries like India where literacy rates are not so high, so people tends to believe in everything they hear as well.

References

[1] Detection of spam-posting accounts on Twitter, *sa Inuwa-Dutse , Mark Liptrott, Ioannis Korkontzelos Department of Computer Science, Edge Hill University, Ormskirk, Lancashire, UK*

[2] Twitter Fake Account Detection , *omputer Engineering Department, Dokuz Eylul University, Izmir, Turkey buketoksuzoglu@iyte.edu.tr ozlem@cs.deu.edu.tr*

[3] Fake Account Detection using Machine Learning and Data Science, *International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-1, November, 2019*

[4] Profile characteristics of fake Twitter accounts, *Big Data Society July–December 2016: 1–13 The Author(s) 2016*

[5] Machine Learning Implementation for Identifying Fake Accounts in Social Network, *nternational Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 4785-4797*

[6] Detection of Fake Profiles on Twitter using Random Forest Deep Convolutional Neural Network, *International Journal of Management, Technology And Engineering*

[7] Fake Account Detection in Twitter Based on Minimum Weighted Feature set, *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering Vol:10, No:1, 2016*

[8] Full Stack Python, *https://www.fullstackpython.com/django.html*