FakeFinder: Social Media Fake Account Detection Using Machine Learning

Geo Joseph

Reg.No: KTE21MCA-2019

Guided By
Dr. Vineetha S
Department of Computer Applications
Rajiv Gandhi Institute Of Technology, Kottayam

March 29, 2023

Contents

Introduction

- Social media platforms are widely used by people at a daily basis
- The number of fake accounts created on these platforms are increasing
- Fake accounts are used to perform scams, impersonation, spreading misinformation etc.
- FakeFinder is a web app that uses machine learning to detect fake accounts
  User has to provide the link to suspicious account

Existing System

- Manually verifying an account is difficult and time consuming
- Rule-based and network-based approaches can be bypassed
- Machine learning based systems provide an average accuracy result on detecting fake accounts

Literature Review

Title	Author	Journal/Conference name	Year of publication	Techniques used	Pros and Cons
Identifying Fake Facebook Profiles Using Data Mining Techniques	Albayati Mohammed, Altamimi Ahmad	Journal of ICT Research and Applications	2019	K-means, K-medoids, Decision Tree, SVM	low accuracy
Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms	Mohammadreza Mohammadrezaei, Mohammad Ebrahim Shiri, Amir Masoud Rahmani1	Security and Communication Networks	2018	Decision trees, SVM, Neural network	High accuracy but limited scope
ke Followers in Twitter: A Machine Learning Approach	Ashraf Khalil, Hassan Hajjdiab, Nabeel Al-Qirim	International Journal of Machine Learning and Computing	2017	Decision trees and random forest	Used large datasets but has limited scope
Prediction of Fake Profiles on Facebook using Supervised Machine Learning Techniques-A Theoretical Model	Suheel Yousuf Wani, Mudasir M Kirmani, Syed mamulAnsarulla	(IJCSIT) International Journal of Computer Science and Information Technologies	2016	SVM, Neural network + Weka tool	Used ensemble classifier but has limited scope

Gap Identified

- FakeFinder provides a public platform for the verification of fake accounts
- Improved accuracy in detecting fake accounts
- Scope of the system includes social media platforms such as Instagram

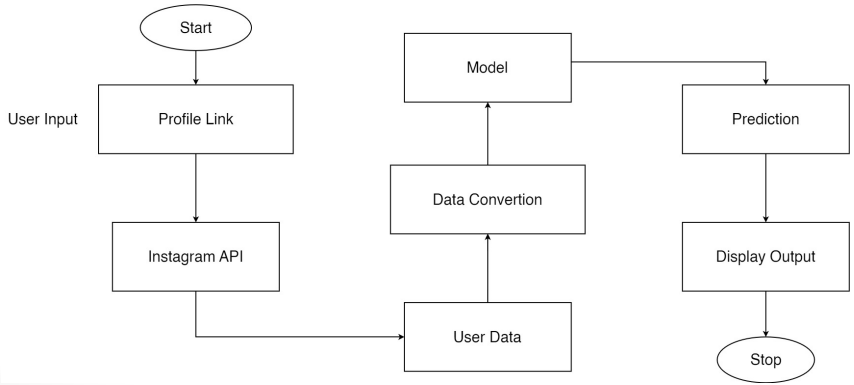
Proposed Work

- The proposed system is a web app that uses machine learning to detect fake accounts.
- Users can input the link of suspicious account
- Application fetches the required data using API
- Returns the prediction result to user
- Uses random forest algorithm

Objectives

- Provide a public platform for the verfication of fake accounts
- Improve the accuracy of fake account detection
- Contribute in improving the credibility and trustworthiness of social media platforms

Design


Start
User Input
Profile Link
Instagram API
User Data
Data Convertion
Model
Prediction
Display Output
Stop

Platform Used

- Operating system: Windows 11
- IDE: PyCharm, Google Colab, VS Code
- Framework: Django
- Python libraries: NumPy, Pandas, SciPy, Scikit-learn, TensorFlow, Keras

Methodology

The process includes :

- Data availability : Evaluating the data that can be accessed through social media APIs
- Data collection and preprocessing : A publically available dataset from Kaggle is used. The attributes are : profile pic, nums/length username, fullname words, nums/length fullname, name==username, description length, external URL, private, posts, followers, follows, fake
- Training the model : Random forest and LSTM are used
- Development : Django is used as the web framework

Screenshots

	precision	recall	f1-score	support
0	0.94	0.98	0.96	105
1	0.98	0.93	0.96	104
accuracy			0.96	209
macro avg	0.96	0.96	0.96	209
weighted avg	0.96	0.96	0.96	209

Fig 1. Classification Report

Screenshots

```
Data :  [[1, 0.18181818181818182, 2, 0.0, 0, 49, 1, 1, 54, 2596, 910]]
   profile pic  nums/length username  fullname words  nums/length fullname  \
0            1                0.181818                2                   0.0

   name==username  description length  external URL  private  #posts  \
0               0                  49             1        1      54

   #followers  #follows
0        2596       910
output:  [0]
geojoseph19 is an Original Account
```

Fig 2. Original account detection

Screenshots

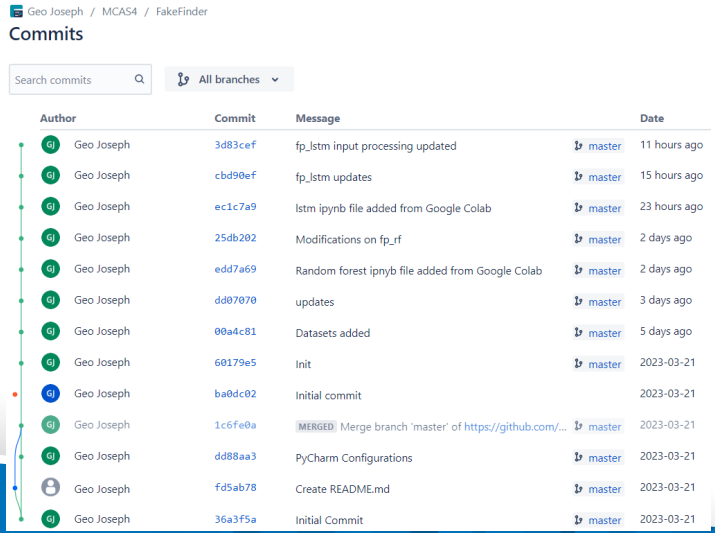
```
Data :  [[1, 0.16666666666666666, 2, 0.0, 0, 0, 0, 0, 1, 11, 6]]
   profile pic  nums/length username  fullname words  nums/length fullname  \
0            1                0.166667               2                   0.0

   name==username  description length  external URL  private  #posts  \
0               0                   0             0        0       1

   #followers  #follows
0          11         6
output:  [1]
mark.zukerburg.324 is a Fake Account
```

Fig 3. Fake account detection

GIT History



Geo Joseph / MCAS4 / FakeFinder

Commits

Search commits

All branches

Author	Commit	Message		Date
Geo Joseph	3d83cef	fp_lstm input processing updated	master	11 hours ago
Geo Joseph	cbd90ef	fp_lstm updates	master	15 hours ago
Geo Joseph	ec1c7a9	lstm ipynb file added from Google Colab	master	23 hours ago
Geo Joseph	25db202	Modifications on fp_rf	master	2 days ago
Geo Joseph	edd7a69	Random forest ipnyb file added from Google Colab	master	2 days ago
Geo Joseph	dd07070	updates	master	3 days ago
Geo Joseph	00a4c81	Datasets added	master	5 days ago
Geo Joseph	60179e5	Init	master	2023-03-21
Geo Joseph	ba0dc02	Initial commit		2023-03-21
Geo Joseph	1c6fe0a	MERGED Merge branch 'master' of https://github.com/...	master	2023-03-21
Geo Joseph	dd88aa3	PyCharm Configurations	master	2023-03-21
Geo Joseph	fd5ab78	Create README.md	master	2023-03-21
Geo Joseph	36a3f5a	Initial Commit	master	2023-03-21

Implementation Status and Plan

No	Tasks	Status	Completion Date
1	Literature Review	Completed	28/02/2023
2	Collecting Data	Completed	13/03/2023
3	Data Preprocessing	Completed	20/03/2023
4	Data Modeling	Completed	24/03/2023
5	Model Validation	Ongoing	-
6	Model Deployment	Not Started	-
7	Web-Application Development	NotStarted	05/04/2023
8	API Connection	Not Started	-
9	Verification	Not Started	-
10	Testing	Not Started	25/04/2023

Thank You!