|| Volume 6 || Issue 4 || April 2021 || ISSN (Online) 2456-0774

INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

DETECTION OF FAKE SOCIAL NETWORK ACCOUNT

Onkar Kadam[1], Nilesh Surse[2]

Department of Computer Engineering Dr. D. Y. Patil School of Engineering Academy, Ambi.[1,2]

--- ***---

This project proposes the use of NLP (Natural Language Processing) techniques to identify "misleading" and "news" reports that originate from unreliable sources. only a count vector (Term Frequency Inverse Document Frequency) (word sizes contrasted to how (word sizes relative to how often they are used in other articles in your dataset) was generated using values from this function vector and those features with relative importance of 1.4 and that feature with relative importance 2.0 (or equal importance) (word) The linguistic models, however, ignore aspects such as word order and meaning. Two documents with completely different word counts can refer to the same thing. As a result, the data science community has put in place various measures to resolve this problem. Facebook is participating in a challenge on Kaggle to remove fabricated news stories from feeds on their social network using AI. fighting fake news is a straightforward job Can you separate "real" and "news" from "fakes?" Thus, the proposed study would have the false and the real news datasets as input, and use the Naive Bayesifier to construct a model that classifies articles by the words they contain. owing to the increased number of online information sources, it is difficult to know what is right and what is incorrect Therefore, the issue of "fake news" has gained further publicity. This research looks at historical and contemporary approaches for determining truth and falsity in text format, as well as how and why it occurs. This paper combines Nave Bayes Classifier, support vector machines, and semantic analysis to identify fake news, coming up with a system of three sections..

Keywords: - *Machine learning, Twitter, fake profiles, online social networks, detection, friends, followers*

--- ***---

I INTRODUCTION

Government propaganda news stories contain propaganda and disinformation, rendering it difficult to write legitimate news and for readers to find accurate data. In our society, false news and media mistrust have reached epidemic proportions. The social media discourse has recently changed from bleating to deceit. Others are now attempting to debunk evidence that contradicts their ideology. The prevalence of misinformation in political discourse in the United States has recently gotten a lot of coverage.

Some have labeled stories that are factually incorrect and misleading as "fake news." The goal of this research is to build a model that can determine whether a given article is true or false. Facebook's image has been widely questioned as a result of media reports. They've already implemented a feature that flags fake news when a user detects it, and they've confirmed that they're working

On an automated system to detect it. That is without a doubt the case. Since fake news can come from both the left and the right, the algorithm must be evenly balanced while also giving both types of sources enough weight. As well as the issue of veracity However, in order to react to this issue, it is necessary to first understand what fake news is. Later on, we'd like to look at natural language processing and machine learning and see if we can spot false news.

In today's Modern society, social media plays a vital role in everyone's life. The general purpose of social media is to keep in touch with friends, sharing news, etc. The number of users in social media is increasing exponentially. Instagram has recently gained immense popularity among social media users. With more than 1 Billion active users, Instagram has become one of the most used social media sites.

Nowadays, Online Social Media is dominating the world in several ways. Day by day the number of users using social media is increasing drastically. The main advantage of online social media is that we can connect to people easily and communicate with them in a better way. This provided a new way of a potential attack, such as fake identity, false information, etc.

A recent survey suggest that the number of accounts present in the social media is much greater than the users using it. This suggest that fake accounts have been increased in the recent years. Online social media providers face difficulty in identifying these fake accounts. The need for identifying these fake accounts is that social media is flooded with false information, advertisements, etc.

After the emergence of Instagram to the social media scenario, people with a good number of followers have been called Social Media Influencers. These social media influencers have now become a go-to place for the business organization to advertise their products and services. The widespread use of social media has become both a boon and a bane for the society. Using Social media for online fraud, spreading False information is increasing at a rapid pace.

Traditional methods cannot distinguish between real and fake accounts efficiently. Improvement in fake account creation made the previous works outdated. The new models created used different approaches such as automatic posts or comments, spreading false information or spam with advertisements to identify fake accounts.

Fake accounts are the major source of false information on social media. Business organizations that invest huge Sum of money on social media influencers must know whether the following gained by that account is organic or not. So, there is a widespread need for a fake account detection tool, which can accurately say whether the account is fake or not.

Due to the increase in the creation of the fake accounts different algorithms with different attributes are use. Previously use algorithms like naive bayes, support vector machine, random forest has become inefficient in finding the fake accounts.

In this paper, we use classification algorithms in machine learning to detect fake accounts. The process of finding a fake account mainly depends on factors such as engagement rate and artificial activity. We came up with an innovative method to identify fake accounts.

We used gradient boosting algorithm with decision tree containing three attributes. Those attributes are spam commenting, artificial activity and engagement rate. We combined Machine learning and Data Science to accurately predict fake accounts.

II LITERATURE SURVEY

STUDY OF RESEARCH PAPER

Paper Name: Detection of Fake and Clone accounts in Twitter using Classification and Distance Measure Algorithms.

Author: Sowmya P and Madhumita Chatterjee.

Abstract: As companies align themselves with individuals and markets, they are referred to as B2B2B2C (OSN). In general, as OSN increases, privacy and data protection levels grow. People who used fake and phished social networking sites were put at risk. reproducing the user identity is a major risk, as it leads to an exact duplication of the current details, and each account threatens that. They can attempt different types of attacks, including phishing, stalking, and mass-mailing. actions that are carried out on social media are the work of a fake identity This paper presents a novel method for identifying fraudulent and counterfeit accounts on Twitter.*The first search method employs a Similarity Comparison algorithm, while the second employs a decision tree.*

Two different forms of knowledge: attribute and network relationship similarity can be used to build a C5.5 decision tree They are compared, to see which one is more powerful.

Paper Name: Machine Learning Implementation for Identifying Fake Accounts in Social Network.

Author: *Rohit Ratur.*

Abstract: *the role of social media in large-scale data dissemination and production of data cannot be understated The volumes of social data will overwhelm even Google's data centers by 2025. Fake accounts are the at an exponential rate, and, and this paper offers a blueprint for locating them on Facebook. In this study, we will use machine learning to predict more precisely the classification of false accounts by determining the strength of their wall posts and post activities. We will use Facebook and Twitter for this purpose, which involves both the use of data protection and authenticity and accessibility.*

Cyberspace is the equivalent of social media "tweets" and "tagging's," which serves to identify and remove fake and harmful content. we use Twitter as our key data source, and sentiment analysis is used to determine how to process the data.

Paper Name: Profile characteristics of fake Twitter accounts.

Author: *Supraja Gurajala, Joshua S White,Brian Hudson, Brian R Voter and Jeanna N Matthews.*

Abstract: *Both the size of a the company or the individual's audience in social networks have a great deal to do with their overall popularity and their social standing. If the number of false accounts on these social media increases, it becomes more difficult to judge the* audience's popularity. More than 62 million accounts have been verified, and a method for automating the identification of robots has been invented.

A fair number of fraudulent accounts have been discovered (well over 1 percent of total users). The difference between these fraudulent profiles and the real profiles is brought to light when the time and URL is scrutinized. The follower data provided a more reliable estimate of the two groups of users.

Fake users had a median number of 30:1, which was in line with previous data, while average users had a ratio of 15:1, which means that indicates that the majority of users were friends. two-year results for ground-based reality users show that the number of friends increased while the number of followers decreased If based on our results, a list-based approach can be used to identify a profile, a short list of active users is a viable.

Paper Name: Detection of Fake Twitter Accounts with Machine Learning Algorithms.

Author: İlhan AYDIN, Mehmet SEVİ, Mehmet Umut SALUR.

Abstract : The many lives of individuals now hang in the balance as a result of social media. Much has already been accomplished in these three fields, including contact, advertising, news, and agenda advancement Misinformation is sometimes used on Twitter, particularly by some malicious accounts.

Social networking is one of the most critical subjects in the business world today. For that reason, it is critical to pinpoint a malicious account. Machine learning methods were applied in this study to try to identify accounts that could be manipulated to look like the real ones The data has been analyzed for this particular purposes, and learning algorithms have been used to identify and delete fake accounts.

Described by means of a decision tree, logistic regression, and a machine learning algorithm When these approaches are compared, the logistic regression outperforms them.

Paper Name: Malicious Account Detection on Twitter Based on Tweet Account Features using Machine Learning.

Author: Farhan Nurdiatama Pakaya, MuhammadOkky Ibrohim,Indra Budi.

Abstract: Twitter faces serious obstacles as a social network as a result of its widespread use. As a consequence, a considerable number of people engaged in online cybercrime. Malicious Internet accounts are present. Spambots and fake followers are examples of fraudulent accounts that might pull down the social networking platform for others.

Spamming bots can be used to send offensive messages to the general public, and the number of followers can be manipulated to give the impression of trust or authority.

Researchers have conducted a number of studies in order to develop a system for detecting malicious accounts that is largely based on graph and profile analysis features.

Malicious and legitimate Twitter accounts can also use Twitter in various ways. Only account information from tweets was used to construct a classification model in this experiment. To separate legitimate accounts from bot accounts, we use additional classifications. Using tfidf features and the XGBoost algorithm, 100 percent accurate malicious or legitimate account detection was achieved, with 95.2 percent on all three types of accounts.

III SYSTEM ANALYSIS

SYSTEM ARCHITECTURE:

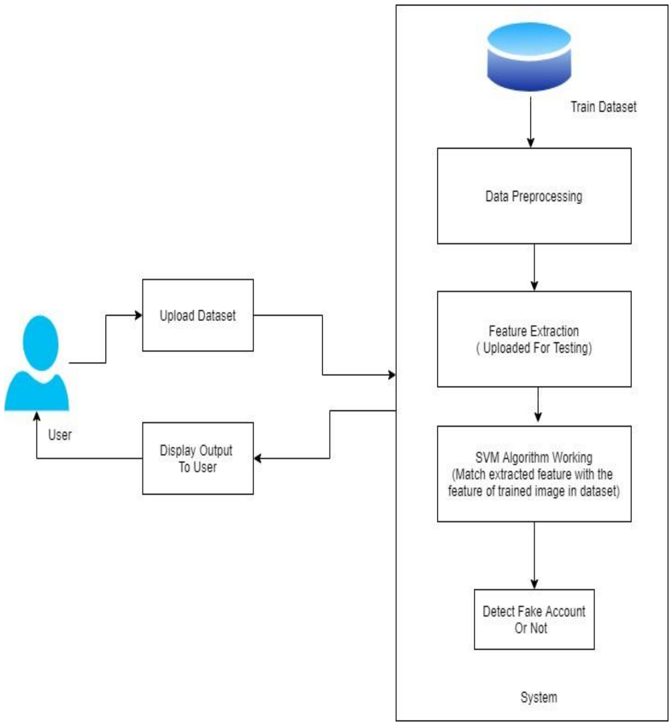


Figure 1: Architecture Diagram

Data Flow Diagrams:

In Data Flow Diagram, we Show that flow of data in our system in DFD0 we show that base DFD in which rectangle present input as well as output and circle show our system, In DFD1 we show actual input and actual output of system input of our system is text or image and output is rumor detected like wise in DFD 2 we present operation of user as well as admin.

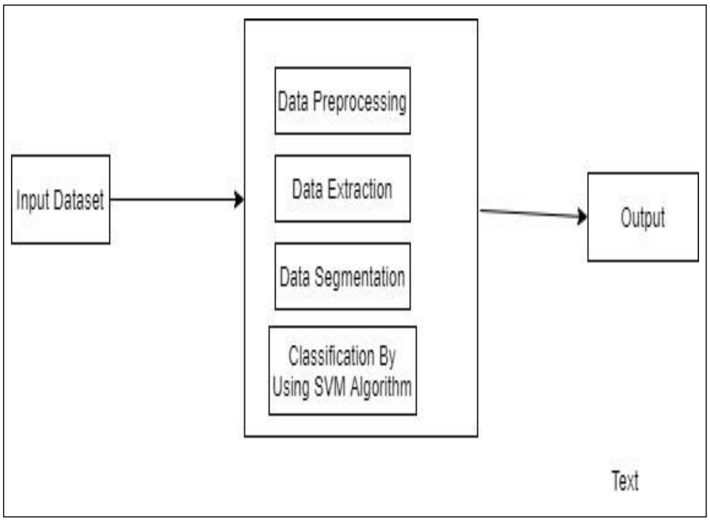
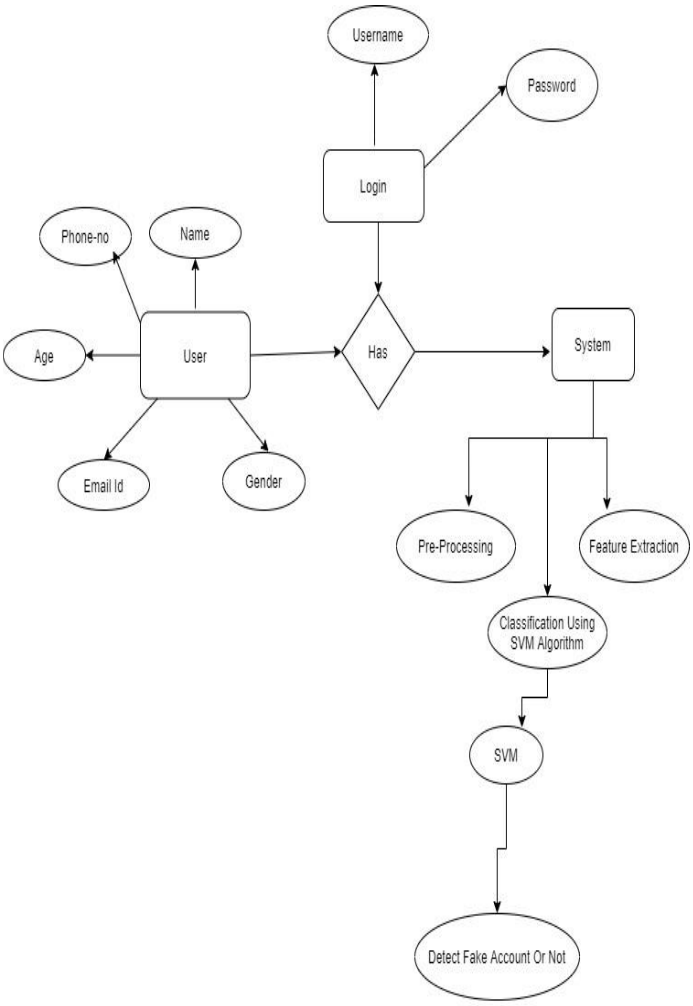


Figure 2: Data Flow Diagram

Entity Relationship Diagram:

ER model stands for the Entity Relationship diagram in our project Dataset, Preprocessing, Train Machine, Machine Learning Algorithm, Build Module and Classification are the entities. Train data and test data are the attributes of Dataset.

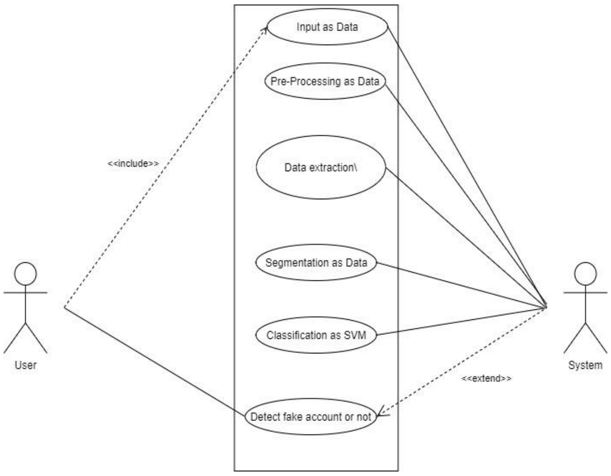
Feature extraction, stop word removal and steaming are attributes of the preprocessing.



Unified Modelling Language is a standard language for writing software blueprints. The UML may be used to visualize, specify, construct and document the artifacts of a software intensive system. UML is process independent, although optimally it should be used in process that is use case driven, architecture-centric, iterative and incremental. The Number of UML Diagram is available

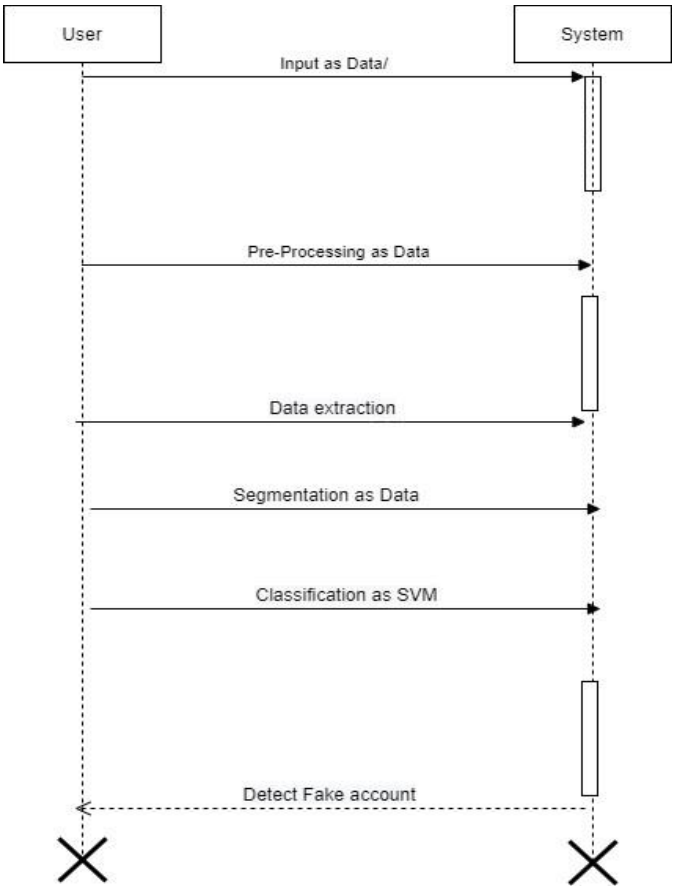
Use Case Diagram:

Use case diagram is used to show which operations are performed by the user and which operation are performed by the system.

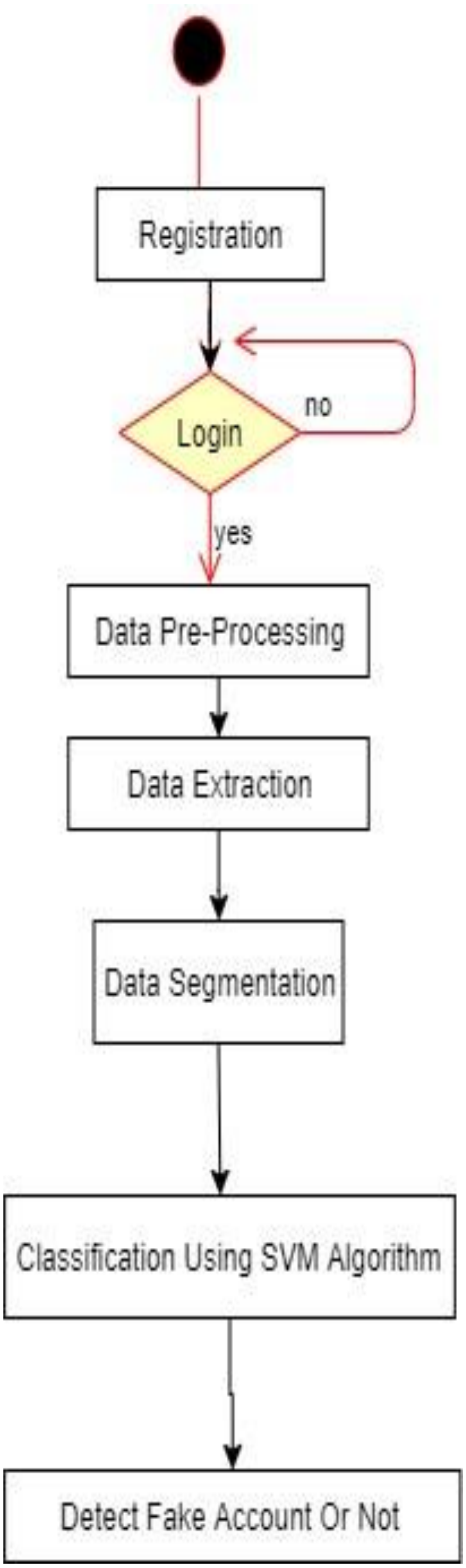


Sequence Diagram:

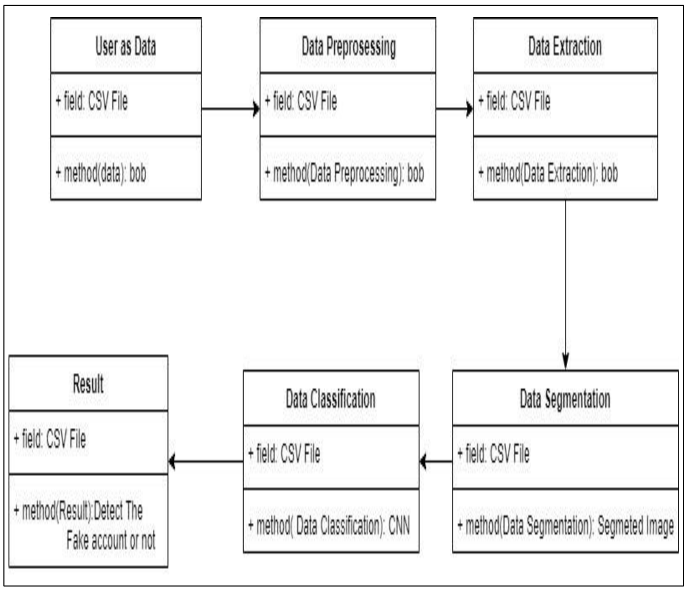
In sequence diagram step by step sequence of steps is shown. In above diagram first preprocess all train data and test data. Then by applying the train data Train the machine and build the module and at the last apply machine learning algorithm on it. For testing purpose apply the test data on module and see the classification either fake or real.



Activity Diagram: Activity Diagram shows the active flow of the system. In above diagram the flow of our project is shown actually how the data flow.



Class Diagram:



IV CONCLUSION

This paper details how to spot fake news on well-known Twitter sources. Additionally, crowdsourced "user-generated “content is more conveniently and inexpensively differentiating between true and false news on Twitter.

In my opinion, social media participants may well find benefits from such a system in growing and endorsing their own reputation. These findings will assist in distinguishing whether social media narratives conform to traditional stories. There is an increase in the number of people who turn to social media for news rather than more conventional outlets.

Conversely, however, social media has been used to propagate fake news and has had a direct impact on both the individual people and the general public. we carried out an in-depth study on the subject of fake news by analyzing literature on both identification and detection. We introduced the fake news definitions and values in both traditional and social media.

REFERENCES

1.Vinod Bharat et al. “Study of Detection of Various types of Cancers by using Deep Learning: A Survey”, International Journal of Advanced Trends in Computer Science and Engineering, 2019, Volume 8 Issue 4,pp 1228-1233

2.Vinod Bharat et al. “A review paper on data mining techniques”, International Journal of Engineering Science and Computing (IJESC), 2016, Volume 6 Issue 5, pp 6268-6271.

3.M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on

Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.

4. Conroy, N., Rubin, V. and Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1), pp.1-4.

5.A. A. Memon, A. Vrij, and R. Bull, Psychology and law: Truthfulness, accuracy and credibility. John Wiley & Sons, 2003.

6.https://www.researchgate.net/publication/270571080_Towards_News_Verification_Deception_Detection_Methods_for_News_Discourse.

7.Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. Technical report,National Bureau of Economic Research, 2017.

8.S. R. Maier, “Accuracy Matters: A Cross-Market Assessment of Newspaper Error and Credibility,” Journalism & Mass Communication Quarterly, vol. 82, no. 3, pp. 533–551, 2005.

9.T. Mitra and E. Gilbert, “CREDBANK: A Large-Scale Social Media Corpus With Asso- ciated Credibility Annotations,” International AAAI Conference on Web and Social Media (ICWSM), 2015. [Online]. Available: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10582.