# SYNOPSIS

## ON

## FAKE NEWS DETECTION (USING MACHINE LEARNING)

**Submitted By:**                                        **Submitted To:**

+ **Name-Pranjal Bansal**                      **Mr. Amir Khan**
  **Sec & Branch-F(CSE)**
  **University Roll no.-191500569**

+ **Name-Sonia Sahu**
  **Sec & Branch-C(CSE)**
  **University Roll no.-191500820**

+ **Name-Shriyanshi Patwa**
  **Sec & Branch-F(CSE)**
  **University Roll no.-191500790**

+ **Name-Ojasva Saxena**
  **Sec & Branch-I(CSE)**
  **University Roll no.-191500520**

# INTRODUCTION

As an increasing amount of our lives is spent interacting online through social media platforms, more and more people tend to hunt out and consume news from social media instead of traditional news organizations. The explanations for this alteration in consumption behaviors are inherent within the nature of those social media platforms: (i) it's often more timely and fewer expensive to consume news on social media compared with traditional journalism , like newspapers or television; and (ii) it's easier to further share, discuss , and discuss the news with friends or other readers on social media. For instance, 62 percent of U.S. adults get news on social media in 2016, while in 2012; only 49 percent reported seeing news on social media . It had been also found that social media now outperforms television because the major news source. Despite the benefits provided by social media, the standard of stories on social media is less than traditional news organizations. However, because it's inexpensive to supply news online and far faster and easier to propagate through social media, large volumes of faux news, i.e., those news articles with intentionally false information, are produced online for a spread of purposes Fake news is typically manipulated by propagandists to convey political messages or influence . Third, fake news changes the way people interpret and answer real news, for instance, some fake news was just created to trigger people's distrust and make them confused; impeding their abilities to differentiate what's true from what's not. It's crucial that we build up methods to automatically detect fake news broadcast on social media. Internet and social media have made the access to the news information much easier and comfortable . Often Internet users can pursue the events of their concern in online form, and increased number of the mobile devices makes this process even easier. But with great possibilities come great challenges. Mass media have an enormous influence on the society, and because it often happens, there's someone who wants to require advantage of this fact. Sometimes to realize some goals mass-media may manipulate the knowledge in several ways. This result in producing of the news articles that isn't completely true or maybe completely false. There even exist many websites that produce fake news almost exclusively. They

intentionally publish hoaxes, half-truths, propaganda and disinformation asserting to be real news – often using social media to drive web traffic and magnify their effect. The most goals of faux news websites are to affect the general public opinion on certain matters (mostly political). Many scientists believe that fake news issue could also be addressed by means of machine learning and AI. The goal of the research is to look at how these particular methods work for this particular problem given a manually labelled news dataset and to support (or not) the thought of using AI for fake news detection. The difference between these article and articles on the similar topics is that during this paper Logistic Regression was specifically used for fake news detection; also, the developed system was tested on a comparatively new data set, which gave a chance to gauge its performance on a recent data.

# EXISTING SYSTEM

There exists a large body of research on the topic of machine learning methods for deception detection, most of it has been focusing on classifying online reviews and publicly available social media posts. Particularly since late 2016 during the American Presidential election, the question of determining 'fake news' has also been the subject of particular attention within the literature.

Conroy, Rubin, and Chen outlines several approaches that seem promising towards the aim of perfectly classify the misleading articles. They note that simple content-related n-grams and shallow parts-of-speech (POS) tagging have proven insufficient for the classification task, often failing to account for important context information. Rather, these methods have been shown useful only in tandem with more complex methods of analysis. Deep Syntax analysis using Probabilistic Context Free Grammars (PCFG) have been shown to be particularly valuable in combination with n-gram methods. Feng, Banerjee, and Choi [2] are able to achieve 85%-91% accuracy in deception related classification tasks using online review corpora.

Feng and Hirst implemented a semantic analysis looking at 'object: descriptor' pairs for contradictions with the text on top of Feng's initial deep syntax model for additional improvement. Rubin, Lukoianova and Tatiana analyze rhetorical structure using a vector space model with similar success. Ciampaglia et al. employ language pattern similarity networks requiring a pre-existing knowledge base.

# <ins>USE OF THE PROJECT</ins>

Increasing use of internet has made it easy to spread the false news. Different social media platforms can be used to spread fake news to a number of persons.With the share option of these platforms, the news spread in a fast way. Fake news just not only affects an individual but it   can also affect an organization or business. So controlling the fake news is

mandatory. A person can know the news is fake only when he knows the complete story of that topic. It is a difficult task because most of the people do not know about the complete story and they just start believing in the fake news without any verification.

The question arises here how to control fake news because a person cannot control the fake news.

The answer is machine learning. Machine learning can help in detecting the fake news. Through the use of machine learning these fake news can be detected easily and automatically. Once someone will post the fake news, machine learning algorithms will check the contents of the post and will detect it as a fake news. Different researchers are trying to find the best machine learning classifier to detect the fake news . Accuracy of the classifier must be considered because if it failed in detecting the fake news then it can be harmful to different persons. The accuracy of the classifier depends on the training of this classifier. A model that is trained in a good way can give more accuracy.

There are different machine learning classifiers are available that can be used for detecting the fake news that will be answered in the next question

# Feasibility of Project

## Technical Feasibility:

The proposed system is developed using Machine Learning .The proposed system will run under Windows 10, core i5 environment. As Windows is very user friendly and GUI OS it is very easy to use. All the required hardware and software are readily available in the market. Hence the system is technically feasible.

## Operational Feasibility:

The proposed system is operationally feasible because of the following reasons:

- The proposed system will aware users regarding false information.
- The cost of the proposed system is almost negligible when compared to the benefits gained.

## Economical Feasibility:

As the necessary hardware and software are available in the market at a low cost, the initial investment is the only cost incurred and does not need any further enhancements. Hence it is economically feasible. The system is feasible in all respects and hence it encourages taking up the system design

# FUNCTIONAL SPECIFICATION

Take a valid news article URL from user. Extract relevant text from the URL, provided by the user, using Scrapy, Beautiful soup. Then we will extract relevant features from the text using NLP (Natural Language Processing). Correctly classify news article as fake news or credible news using different machine learning models/algorithms.

Each user can view all the recently processed and classified news articles and verify the correctness of the classification by voting. After a predefined limit of time and number of votes we can verify that whether the software classified a given news article correctly or not. We can then modify our classification if needed and add the news article in the training set to improve accuracy of future predictions

# Software Specification

- Technology Implemented    -    Machine Learning
- Language Used    -    Python 3.2
- Database    -    SQL
- User Interface Design    -    Graphical User Interface
- Web Browser    -    Google Chrome

# Hardware Requirements

- Processor    -    i3 or above
- Operating System    –    windows, mac, linux
- RAM    –    2GB (atleast)
- Hardware Devices    –    Internet supportive pc with basic requirements
- Hard Disk    -    62GB
- Display    –    LCD, OLED, AMOLED and others

# FUTURE SCOPE

The fake news challenge is perilous and is spreading rapidly like a wildfire as it becomes easier for information to reach the mass in various flavor. Reports have shown that, just like in the last US presidential elections, fake news can have a huge impact in politics and thereafter on the people like a domino effect. With the help of artificial intelligence, we can control and limit the spread of such misinformation more quickly and efficiently as compared to manual efforts. The work in this project proposes a stacked model which fine tunes the informational insight gained from the data at each step and then tries to make a prediction. Although many attempts have been made to solve the problem of fake news, any significant success is yet to be seen. With huge amounts of data collected from social media websites like Facebook, Twitter, etc., the best models improve every day. With the use of deep neural networks, the future work in this field seems a lot more promising. The limitations that come packaged with this problem is that, the data is erratic and this means that any type of prediction model can have anomalies and can make mistakes. For future improvements, concepts like POS tagging, word2vec and topic modelling can be utilized. These will give the model a lot more depth in terms of feature extraction and fine-tuned classification.