Fake News Detection

Project Guide Name:- Dr. Rahul Ambekar

Group Members:	Moodle ID:
1. Shubham Padte	16102058
2. Pankit Khimasiya	16102034
3. Ashish Kothari	16102045

Abstract :

Social media for news consumption is a double-edged sword. A large body of recent works has focused on understanding and detecting fake news stories that are disseminated on social media. To accomplish this goal, these works explore several types of features extracted from news stories, including source and posts from social media. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society. First, fake news is intentionally written to mislead readers to believe false information, which makes it difficult and nontrivial to detect based on news content; therefore, we need to include auxiliary information, such as user social engagements on social media, to help make a determination. The proposed application will be implemented to verify the gentility of news. We believe that interesting findings on the usefulness and importance of features for detecting false news will prevent readers from misleading. This project will filter the news by using Natural language processing and Machine learning (supervised learning), by processing large dataset of news. Finally, we discuss how fake news detection approaches can be used in the practice, highlighting challenges and opportunities.

Introduction:

As an increasing amount of our lives is spent interacting online through social media platforms and other online platforms , more and more people tend to seek out and consume news from social media rather than traditional news organizations. The reasons for this change in consumption behaviours are inherent in the nature of these social media platforms: (i) it is often more timely and less expensive to consume news on social media compared with traditional news media, such as newspapers or television; and (ii) it is easier to further share, comment on , and discuss the news with friends or the other reader on social media. Our economies are not immune to the spread of fake news either, with fake news being connected to stock market fluctuations and massive trades. For example, fake news claiming that Barack Obama was injured in an explosion wiped out $130 billion in stock value

Objectives:
The major objective is detect fake news , which is a classic text classification problem with a straight forward proposition. It is needed to build a model that can classify " fake" or "real" news. To deliver user our opinion on news and make them classify the news is fake or real. To make user easier to differentiate news.

Literature Review:-
There are several approaches to tackle the problem of unreliable content on social media. Some authors opt by analysing the patterns of propagation, others by creating supervised systems to classify unreliable content, and others by focusing on the characteristics of the accounts that share this type of content. In addition, some works also focus on developing techniques for fact-checking claims or focus on specific case studies. Some paper characterised on four perspectives (i) knowledge-based (ii) style-based (iii) propagation-based (iv) credibility-based

Problem Definition:-
In today's days social media for news is dangerous. A large body of recent works has focused on understanding and detecting fake news stories that are disseminated on social media. To accomplish this goal, these works explore several types of features extracted from news stories, including source and posts from social media. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society. The proposed application will be implemented to verify the gentility of news.

Scope:

A model will be trained by letting it read millions of documents. These documents included journalism and scientific journal articles, satire articles, narrative fiction, opinion pieces, blogs, politically leaning news and even hate speech examples.

The model still learning all the time. If you think the model's opinion wasn't quite right then we will include that feedback as a training dataset for the model which will continue to learn and develop using the new data. The model will never tell you if an article is True or Not. That's not the model's job.

Our philosophy is that the best judge of Truth from Fiction is you. Model's job is to read what you give us carefully and to let you know what kind of article it is. Is the article biased or does it give a measured assessment? Is the article written in a way that focuses on facts, or is it written with an agenda? Does the writing style use manipulative persuasion or is written in a way that provides you with the information to draw your own conclusions? Our model will make it easier to sort out the truth yourself.

Along with learning which articles correspond to which labels, the model also learns which part of the article to pay the most attention to in context with the other sentences. If Model thinks that the articles shows high Opinion, Sensational, Agenda, or Satire, then you can see the sentences that most influenced these labels here.

Label Definitions

- **Journalism**
    - Journalism focuses on sharing information. These articles do not attempt to persuade or influence the reader by means other than presentation of facts. Journalistic articles avoid opinionated, sensational or suspect commentary. **Good journalism does not draw conclusions for the reader** unless manifestly supported by the presented evidence. Journalistic articles can make mistakes, including reporting statements that are later discovered to be false, however the mark of good journalism is responsiveness to new information (via a follow up articles or a retraction) especially if it counters previously reported claims.
- **Wiki**
    - Like Journalism, the primary purpose of Wiki articles is to inform the reader. Wiki articles do not attempt to persuade or influence reader by means other than presentation of the facts. **Wiki articles tend to be pedagogical or encyclopaedic in nature**, focusing on scientific evidence and known or well-studied content, and will highlight when a claim or an interpretation is controversial or under dispute. Like Journalism, Wiki articles are responsive to new information and will correct or retract prior claims when new evidence is available.
- **Satire**
    - **Satirical articles are characteristically humorous, leveraging exaggeration, absurdity, or irony** often intending to critique or ridicule a target. Claims in works of satire may be intentionally false or misleading, tacitly presupposing the use of exaggeration or absurdity as a rhetorical technique. Satirical articles can often be written in a journalistic voice or style for humorous intent.
- **Sensational**
    - Sensational articles provoke public interest or excitement in a given subject matter. **Sensational articles tend to leveraging emotionally charged language, imagery, or characterizations** to achieve this goal. While sensational articles do not necessarily make false claims, informing the reader is not the primary goal, and the presentation of claims made in sensational articles can often be at the expense of accuracy.
- **Opinion**
    - Opinion pieces present the author's judgments about a particular subject matter that are not necessarily based on facts or evidence. Opinion pieces can be written in a journalistic style (as in "Op-Ed" sections of news publications). **Claims made in opinion pieces may not be verifiable by evidence or may draw conclusions that are not materially supported by the available facts.** Opinion pieces may or may not be political in nature, but often advocate for a particular position on a controversial topic or polarized debate.
- **Agenda-driven**
    - **Agenda-driven articles are primarily written with the intent to persuade, influence, or manipulate the reader to adopt certain conclusions.** Agenda-driven articles may or may not be malicious in nature, but characteristically do not convince the reader by means of fact based argumentation or a neutral presentation of evidence. Agenda-driven articles tend to be less reliable or more suspect than fact based journalism, and an author of agenda-driven material may be less responsive to making corrections or drawing different conclusions when presented with new information.

Technology Stack:

- Multiple regression techniques will be used like logistic regression, ridge regression, lasso regression and the most accurate technique will be chosen.
- Extensive use of neural networks will be done.
- The training and testing will be done using python language as it is effective and easy to implement
- The ide used will be jupyter notebook which is a great tool for data analysis.
- The jupyter notebook is available in anaconda, also spyder will be used for some data visualization which is also available in the same

Benefits for society:

Correcting the fake news is one of the most important aspects for the society. Main goal of this project is to prevent the society for being misguided as well as spreading the rumours. By reducing the spread of rumours there can be prevention of chaos in people.

Application:

Fake news destroys your credibility. If your arguments are built on bad information, it will be much more difficult for people to believe you in the future.

Fake news can hurt. Purveyors of fake and misleading medical advice like Mercola.com and NaturalNews.com help perpetuate myths like HIV and AIDS aren't related, or that vaccines cause autism. These sites are heavily visited and their lies are dangerous.