MINI PROJECT (2020-21)

ON

Misinformation: Detecting and CombATING MISINFORMATION ON SOCIAL MEDIA .

**AUTHOR:** SANA AFREEN

**ABSTRACT:**

The widespread of misinformation in social media has become a issue.

As in this survey,we aim to consolidate the observations and investigate how an optimal method can be selected by given specific conditions and contexts.

**ABOUT DATA:**

Data contains the basic information(Title,Text,Subject,Date)

**PROBLEM STATEMENT:**

Write a program to detect and combat misinformation on social media.

It should have following features:-

1. Using Python.

1. We are making this program by using python,jupyter notebook and by using suitable

Libraries.

.

3) Taking two CSV file one is for fake news and other one is for True news.

4) Finding those news resources , that is where they came from.

5) Detecting the news whether they are fake or real by given information.

6) Using two functions :Logistic Regression And Decision tree classification for finding

accuracy of our model.

7) By using this, we can find fake and real news.

**INTRODUCTION:**

The openness and timeliness of social media have largely facilitated the creation and dissemination of misinformation, such as rumor, spam, and fake news. As witnessed in recent incidents of misinformation, how to detect misinformation in social media has become an important problem. By definition, misinformation is false or inaccurate information that is deliberately created and is intentionally or unintentionally propagated.

**MISINFORMATION** : There are several related terms similar to misinformation. Rather than the concepts are relatively easier to distinguish, such as spam (a large number of recipients) rumor (verified or unverified) and fake news (in the format of news).

**Misinformation detection** seems to be a classification problem, which has the same setting as text categorization.

***• Content-based misinformation detection:*** directly detecting misinformation based on its content, such as text, images and video.

***• Context-based misinformation detection:*** detecting misinformation based on the contextual information available in social media, such as locations and time.

***• Propagation-based misinformation detection:*** detecting misinformation based on the propagation patterns, i.e., how misinformation circulates among users.

***• Early detection of misinformation:*** detecting misinformation in an early stage before it becomes viral, usually without adequate data or accurate labels.

**REQUIREMENTS:**

**LANGUAGE USED**: Python

**HARDWARE REQUIREMENTS**: Windows Desktop

**SOFTWARE REQUIREMENTS**: Jupyter Notebook

**REQUIRED INSTALLMENTS:**

**-**Python

**-**Jupyter

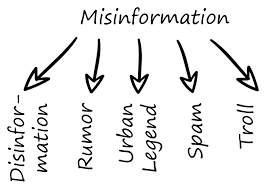
**LIBRARIES:**

**-**Pandas

**-**Matplotlib

**-**Numpy

**-**Sklearn



**• How are misinformation spreaders spreading misinformation and attracting attention?**

Existing research mostly focuses on the spreaders - or the accounts that post misinformation in social media. In the real world, a spreader would more than that to “spread” misinformation, such as commenting under certain topics, making friends with similar communities, and even privately messaging interested accounts. In addition to detecting them, it would be interesting to discover and understand such spreading behaviors, which may ultimately facilitate building a robust detection system.

**• How to make detection methods robust to adversarial attacks, or how to exploit adversarial learning to enhance a detection method?**

Adversarial machine learning aims to enable machine learning methods to be robust andeffective in the presence of adversarial attacks. Current research focuses on adversarialattacks of misinformation spreaders, however, if there is a malicious adversary that haspartial or full knowledge of the misinformation detection algorithm, existing methods can be vulnerable. It will be interesting to discover robust methods in the presence ofadversarial attacks.

**ALGORITHM:**

1-START

2-Make necessary imports.

3-Now,lets read the data for both fake news and true news , then get the shape of

the data and the first five records for both true and false dataset.

4-Now take rows of each dataset for manual testing.

5-Save the dataset as CSV file , but before saving lets merge these data cell into one.

6-Merging main dataset that are dffake and dftrue.

7-Now removing , title, subject , date , from main data.

8-Now shuffling dataset.

9-Checking null value.

10-Removing all special characters/links.

11-Convert data text.

12-Splitting dataset into train and test.

13-Vectorizing that x variable by using TfidfVectorizer.

14-Using Logistic Regression and Decision Tree Classification for checking accuracy

of our model.

By using Logistic Regression: 98.7%

By using Decision Tree Classification: 99.5%

15-STOP

**CONCLUSIONS:**

This project that I have projected in a system will help for detecting and combating Misinformation on social media.

By this project,we can find real and fake news which are spreading all over

the internet.

**REFERENCES:**

* S. Kwon and M. Cha. Modeling bursty temporal pattern of rumors. In ICWSM, 2014. [20] K. Lee, J. Caverlee, and S. Webb. U
* Research papers about misinformation on social media.
* http://users.wpi.edu