# **Project Title:** **Detecting Fake News**

# **Student Name:** Rahil Deepakkuamr Pancholi

# **Date:** 20-03-2023

## **Project Overview:** Provide a brief description of your project, including the objectives and scope.

Identifying and evaluating incorrect or misleading material presented as if it where true. The goals of identifying fake news are to stop misinformation from spreading, to protect individuals from harm caused by incorrect information, and to promote accuracy and honesty in news reporting. This collection of data, fake and true articles is gathered from the website of the University of Victoria.

The scope of detecting fake news

Fact-checking involves reviewing the sources mentioned and verifying the truth of assertions made in news reports. Source evaluation involves deciding the reliability of the information sources referenced in news reports.

The model will be trained on a large dataset of false and authentic news items using well-known Python frameworks like as TensorFlow, Keras, or PyTorch. Certain aspects of the text will be scrutinised. Following that, a machine learning model will be built using these attributes to decide if the news is true or false.

## **Goals and Objectives:** Restate the goals and objectives of your project and assess how well you achieved them. Be specific and provide evidence to support your assessment.

One of the most basic reasons for detecting fake news is to prevent the spread of misleading information that might harm people, groups, or society as a whole.

False news can be used to slander or harm persons or companies' reputations. Identifying false news can aid in the protection of these institutions.

As, we have timeline of what to do in the project based on the time we spent. Till now we have done many tasks like selecting the project, finding the dataset, data assessment as assessing the data is necessary because our project is based on that.

Then the model planning phase come, in which we have done cleaning the data, we checked that does the data have noise or not, in which we found that the data had some blanks and we removed it. Then some data was miss matched so we removed them, while the data that we removed were not much.

Then analysing datasets with Tableau, we got to know about the categories of the news that we have and we find that there are government related news, middle east news, politics and more.

We have created the word cloud that help us to understand more on the data sets that we have.

Our goal is to create a frontend tool which identify weather the given news is real or fake and showing the percentage of that, how many percent the model giving to the news with probability of being true or fake. We have **created the frontend Home page which is easy to use and understand**.

We will going to use Machine learning and NLP system that can identify if the news is real or fake.

While performing some models in python, using decision tree classifier the model predict score that we got is 0.9955.

## **Methodology:** Describe the methodology you used to complete your project, including any research methods or tools you used. Assess the effectiveness of your process and whether any changes were necessary.

Data collection: We took the data sets from University of Victoria website.

Data pre-processing: involves cleaning the data by eliminating any irrelevant or redundant information and turning it into an easily analysed format.

EDA: We carried out EDA on the dataset using Tableau and Python, including the creation of word clouds to identify the most frequently occurring words in True and fake file.

We will use Machine learning models to check the accuracy of datasets, and then we will changes parameters to increase the reliability of the model, so it can delivery higher accuracy.

For that we have used tools like Excel in which the data is available as csv file, then the tubule to create the charts, python to create word clouds, and understand the nuance of the data with the categories of it. Python to work on Models.

## **Results:** Describe the results of your project, including any key findings or insights you gained. Assess the significance of your results and their contribution to the field.

We reviewed the data and there were some blanks, so, we decided to eliminate them. Then we deleted a small number of data that were incorrectly matched, even though there weren't many.

When we learned more about the different types of news that we have, we discovered that there are news stories regarding the government, the Middle East, politics, and more.

In order to better understand the data sets we have constructed a word cloud.

In true file, the subjects were political news and world news.

Which were 11272 and 10154 respectively. And it was recorded in year 2016 and 2017.

Using the word cloud in python, we found that in True file Reuters news, which was the key word of keep pooping up. While candidates name like Hilary and Trump also were the part of the many articles.

While performing some models in python, using decision tree classifier the model predict score that we got is 0.9955.

## **Reflection:** Reflect on your overall experience with the project. What were some of the challenges you faced? What did you learn? How did you grow because of this project?

Till now it was more of seeing the picture from far, like understanding the subjects, what the problem the world faces due to fake news or propaganda of the news.

Then EDA part came in which to check the data sets and then seeing the article from the file which separates them from fake and true. The challenges that faced was to extracting year from the date the published the article, because many were in different format and we have around 40000 rows if we combine the two datasets. But then after some research solve the problem.

Working in team for the capstone project, make sure that putting all files on online platform so everyone can access that and work on the centralize the platform.

The model that we are working on it requires more testing and at the same time searching for other models which are faster and more accurate.

## **Conclusion:** Summarize your project and its outcomes and assess its success. What impact do you think your project will have on the field or community it serves?

Our project is still going on, so our work till now is satisfying and we have more understanding of the datasets and what we want to achieve with it. And the success of it will more depend on the models that we will create and how accurately the model assess the articles. For healthy society we need more fake news detection platform as it will make sure that people will not pursue by the wrong or fake narrative. So, our project will identify truth in articles and create the positive impact on community.

## **Future:** Work Identify any areas for future work or research related to your project.

Future work will be more the evaluation of model, and deployment of model to predict the accuracy of the news.

Then for front end we are working on flask, stream lit and other method for the good representation of our front end.

Training the model: To determine whether news stories are authentic or not, train a machine learning model using the retrieved features, such as a classification algorithm.

Evaluation of the model: Using a validation set, assess the model's accuracy, and make any necessary modifications to enhance its performance.

Implementation: Implement the model in a platform or system that can be used to categorise news stories as true or fake automatically.

## **Acknowledgments:** Acknowledge any individuals or organizations that contributed to your project.

Our professor, because in every session he give us good ideas and feedback on our project and what we have done till now, then letting us know that why the centralize platform is necessary in the working of project and all.

Then search engine google, many articles from google that added the more knowledge and background of fake news detection.

<https://onlineacademiccommunity.uvic.ca/isot/2022/11/27/fake-news-detection-datasets/>