# **Project Title:** [Fake News Detection]

# **Student Name:** [Siddharth Manish Mahidhar]

# **Date:** 03-21-2023

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

The project's main goal is to create a Python-based system that can quickly and easily identify fake news posts on social media sites. The goal of the project is to create and implement a machine learning model in Python that can analyse and classify news stories as legitimate or fraudulent based on a specified set of criteria. These features can contain any type of information, including text, images, video, and other multimedia elements. The model will be trained on a big dataset of fictitious and actual news items using well-known Python libraries like TensorFlow, Keras, or PyTorch.

## **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.

The creation of a machine learning and NLP system that can accurately identify fake news is the main goal of this project. The project's objectives were collecting and preparing data, developing and training machine learning models, evaluating the performance of the models, and developing a user-friendly system interface (Website).

To accomplish the first objective, news article datasets were gathered, cleaned up using text cleaning and tokenization, and then the data was examined using Tableau. Then, we moved closer to executing EDA on it. The most commonly used words in the dataset were visualised using Tableau and created as word clouds in EDA to help us understand the data.

## **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.

The technique for the project covers a number of steps, such as data collection, preprocessing, exploratory data analysis (EDA), and the development and evaluation of machine learning models. This is a detailed description of all the techniques used:

1. Data collection: After searching through a number of websites for the pertinent dataset, we chose one from the University of Victoria website that contains both authentic and fraudulent news stories.

2. Data cleaning and preparation: We removed redundant entries, added or deleted missing values, and changed the text's case to clean up and prepare the data.

3. EDA: We carried out EDA on the dataset using Tableau and Python, including the creation of word clouds to identify the most frequently occurring phrases in real and fake news. Also, we used Python to analyse the datasets by building a few charts. After that, we combined the two datasets into one variable and assigned them the labels "real" and "fake" by adding a new column for this label.

4. Machine Learning Model: To assess the correctness of the dataset and determine whether it is overfitting or underfitting, we constructed machine learning models. We will employ a variety of techniques, such as feature extraction, feature engineering, and model selection, to improve the performance of the models.

5. Evaluation: We will assess the performance of the model using various measures, including accuracy, precision, recall, and F1 score. We will use techniques like cross-validation and confusion matrices to confirm the models' efficacy if necessary.

## **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.

Via EDA, we gained knowledge of the distinctions between true and misleading news reports. For instance, we discovered that the most often utilised words in fake news articles were "Trump," "Hillary," "Obama," and "election." This shows that political figures and events are regularly used in fake news articles to grab readers' attention and sway them.

Given that the dissemination of fake news has grown to be a significant problem in recent years, this project's contribution to the subject is essential. The creation of effective machine learning models for false news identification can help with the detection and prevention of the spread of misleading information. It might be advantageous for society if news sources become more trustworthy and authentic.

## **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?

Having no coding background except the things that I have learnt in the previous two semesters, I had a lot to learn about Machine Learning and also implementing the same in our project. Initially I faced a bit of issues coping with the coding aspect of the project. But, with the help of my team members I was able to learn and understand more effectively. Also working in a group helped me a finish the work with deadlines and have a proactive approach towards the hurdles that came on the way.

## **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?

The detection of fake news has grown in significance and importance over the past few years. The proliferation of social media and online news sources has made it more difficult to distinguish between legitimate news and false information. This problem has the potential to damage both people's personal and social reputations, in addition to political unpredictability and false information about public health.

Using a range of techniques, such as deep learning, machine learning, and natural language processing, we will look for patterns and features that distinguish between true and false news.

The effectiveness of these methods depends on a number of elements, such as the complexity of the algorithms, the quality and variety of the training data, and the algorithms capacity to adapt to new varieties of fake news.

It may have a substantial impact on the field and community that fake news detection serves. By providing trustworthy and accurate information, it can help stop the spread of false information and promote critical thinking and well-informed decision-making. Additionally, it can defend vulnerable populations from damage and assist avoid the inappropriate use of information for commercial or political objectives.

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

1. Enhance algorithm and model: While current techniques for identifying fake news are beneficial, there is still need to enhance models for greater precision. The capacity to differentiate between authentic and fake news more precisely might be substantially boosted by constructing more complex advanced algorithms and models using Python.

2. Multimodal approach: Many fabricated news stories are disseminated on social media platforms, where they may be accompanied by photoshopped or phoney videos. By recognising more layers, a multimodal technique that takes account text and visual features may be able to detect bogus news more successfully.

3. Real Time Detection Extension: Due to how quickly bogus news may spread, real-time detection is necessary to prevent it. The impact of fake news could be greatly minimised by building real time fake news detection systems that can immediately recognise and label fake news as they arrive.

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

I would like to point out that the author of our dataset and various python modules and packages from machine learning and natural language processing area are assisting us in the development of a powerful false news detecting system. Also, I would like to acknowledge our professor Mr. Manoop Paliot for guiding us through out our project.