# **Project Title:** Detecting Fake News

# **Student Name:** Jalpa Tank

# **Date:** 21st March 2023

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

With the rapid increase in the availability and spread of news on social media and the internet, fake news has become a significant problem in the modern world. The goal of this project is to develop a machine learning-based system that can **accurately identify fake news**.

This project covers data collection, preprocessing, feature engineering, and model training in python. A **dataset** of a fake and real articles is collected from a **University of Victoria’s website**. After that, the data from this dataset will be preprocessed and put into a format that machine learning algorithms can understand and process. The text will be examined for a number of factors. Afterwards, a machine learning model will be trained using these features to determine if the news is genuine or fake. The model will be trained using well-known Python libraries like TensorFlow, Keras, or PyTorch on a sizable dataset of fake and real news stories.

The performance of the machine learning model on a test dataset will be used to evaluate the project. We will evaluate the model with the metrics such as accuracy, precision, recall, and F1-score**. The final result** of this project will be a **user-friendly interface** that can receive news/articles as an input and provide a judgement of their reliability that either it’s fake or real. Many of the bad aspects of social media applications can be reduced with the aid of this initiative; we can **prevent the spread of the fake news**.

## **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 primary objective of this project is to create a machine learning and NLP system that can reliably identify fake news. The projects goals included gathering and preparing data, creating and training machine learning models, assessing model performance, and creating a system interface (Website) that is user-friendly.

To meet the first goal, a dataset of news items was collected and preprocessed using text cleaning, tokenization and analyzing the data with the Tableau. We then made a progress towards performing EDA on it. We used Tableau and creation of word clouds in EDA to visualize the most frequently used words in the dataset for gaining insights to the data which can be found here: [**EDA File**](https://drive.google.com/file/d/1Lj92p-WM06udJISbcbLkbzKIyKpB2t7b/view?usp=share_link)

Next, I have created python models one using Logistic Regression and other one is Decision Tree Classifier which has the accuracy of 0.9873 and 0.9955 which are overfitting to the training data, the Model file : [**Model Creation**](https://drive.google.com/file/d/1Lj92p-WM06udJISbcbLkbzKIyKpB2t7b/view?usp=share_link)

Furthermore, I have made a webpage , A home page for our site using Flask, you can find the webpage image link here : [**Home Page**](https://drive.google.com/file/d/1QLHjQzaZiV2dSvSO61TusZtGnaepaGqZ/view?usp=share_link)

## **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 projects methodology includes a number of processes, from gathering the data, preprocessing data, exploratory data analysis (EDA), and the creation and assessment of machine learning models. The methods utilised is described in full below:

1. Data Gathering: After researching various websites for the relevant dataset we finalized one dataset from the University of Victoria website in which real and fake news articles can be found.
2. Preprocessing: We cleaned and prepared the data by deleting duplicate entries, filling or deleting missing values, as well as by changing the text's case.
3. EDA: Using Tableau and python, we performed EDA on the dataset, including the construction of word clouds to pinpoint the most frequently occurring words in real and fake news. Also, we did the analysis on the dataset using python about the datasets by creating few charts and then we combined both datasets in one variable giving them labels as real – 1 and fake – 0 by adding the new column for this label.
4. Machine Learning Model: We created machine learning models to check the accuracy of the dataset and see whether it is overfitting or underfitting. To enhance the performance of the models, we will use number of methods, including feature extraction, feature engineering, and model selection.
5. Evaluation: Using different metrics like accuracy, precision, recall, and F1 score, we will test the models performance. To verify the effectiveness of the models, we will apply approaches like cross-validation and confusion matrices 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.

We learned about the differences between actual and false news stories through EDA. For example, we found that "Trump," "Hillary," "Obama," and "election" were the words most commonly used in false news pieces. This indicates that fake news pieces frequently use political persons and events to attract readers attention and influence them.

This project's contribution to the subject is crucial, given the spreading of fake news has become a major issue in recent years. The detection and prevention of the transmission of misleading information can be improved by the development of efficient machine learning models for fake news detection. By increasing the authenticity and reliability of news sources could benefit society.

As a result you can find the EDA file and the webpage below in the link, also the model accuracy for is Logistic Regression – 0. .9873 and for Decision Tree Classifier is 0.9955 which is overfitting to the training data so that we have to improve our model with more advancement and complex features.

EDA FILE : [**https://drive.google.com/file/d/1Lj92p-WM06udJISbcbLkbzKIyKpB2t7b/view?usp=share\_link**](https://drive.google.com/file/d/1Lj92p-WM06udJISbcbLkbzKIyKpB2t7b/view?usp=share_link)

Webpage: [**https://drive.google.com/file/d/1QLHjQzaZiV2dSvSO61TusZtGnaepaGqZ/view?usp=share\_link**](https://drive.google.com/file/d/1QLHjQzaZiV2dSvSO61TusZtGnaepaGqZ/view?usp=share_link)

## **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?

Detecting fake news can be a difficult and crucial part. I've had past IT experience, with technical knowledge and few python projects may have helped me approach the assignment more efficiently.

Finding trustworthy sources of information and separating them from unreliable sources is one of the primary challenges in this project. In order to classify the facts, this requires thorough investigation and examination of various news sources.

The models robustness and ability to function well over a range of various datasets and scenarios can be a challenge as well. To identify the most effective combination for spotting fake news, this may need significant testing and experimenting with various models and data sets.

Reflecting on my total experience, I can say that I have learnt a lot about the complexity of machine learning and natural language processing techniques. In addition, I have also learned about data analysis methods, testing, and experimenting that I may apply to future projects.

Overall, the fake news detection project gave me the chance to develop new skills and expand my knowledge in this area.

## **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?

In recent years, the detection of fake news has become a significant subject of increasing importance. It has become more hard to differentiate between real news and fake news as a result of the growth of social media and online news sources. In addition to political instability and public health misinformation, this issue has the potential to harm both people's individual and social reputations.

We will search for patterns and features that distinguish between real and fake news using a variety of methods, including deep learning, machine learning, and natural language processing.

The effectiveness of these methods depends on a number of variables, such as the compl

exity of the algorithms, the quality and variety of the training data, and the algorithms capacity to adapt to new varieties of fake news.

The field and community that fake news detection serves may be significantly impacted. It can reduce the spread of incorrect information and encourage critical thinking and well-informed decision-making by offering reliable and accurate information. Additionally, it can safeguard vulnerable groups from harm and help stop the inappropriate use of information for commercial or political ends.

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

1. Improve algorithm and model : Current methods for detecting fake news is helpful, but there is still chance to improve models for better accuracy. The ability to identify between real and fake news more accurately might be greatly increased by creating more sophisticated advanced algorithms and models using Python.
2. Multimodal approach: Many fake news articles are spread via social media sites, where they may be supported by false or edited photographs or videos. By identifying more layers, a multimodal approach that takes into text and visual elements may be able to detect fake news more effectively.

## **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.